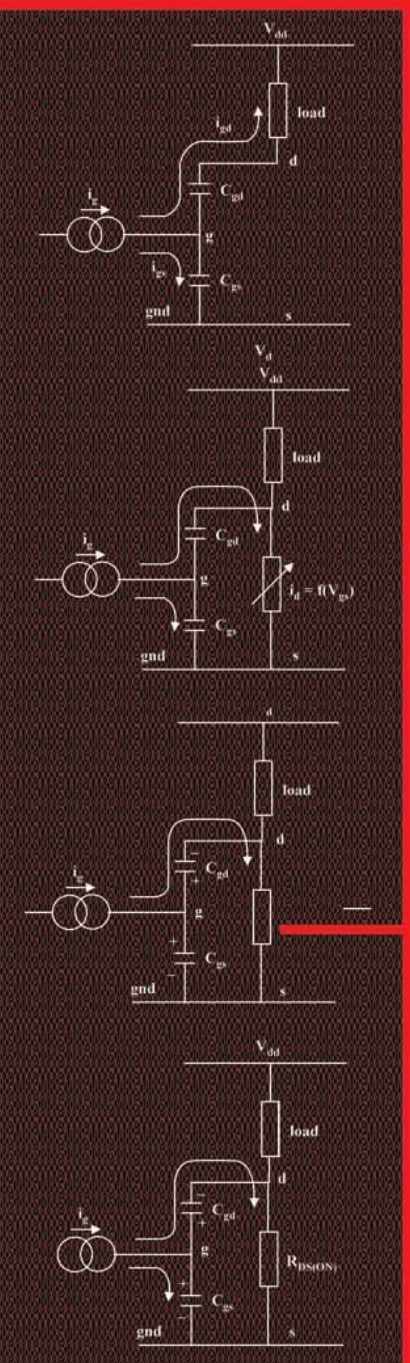
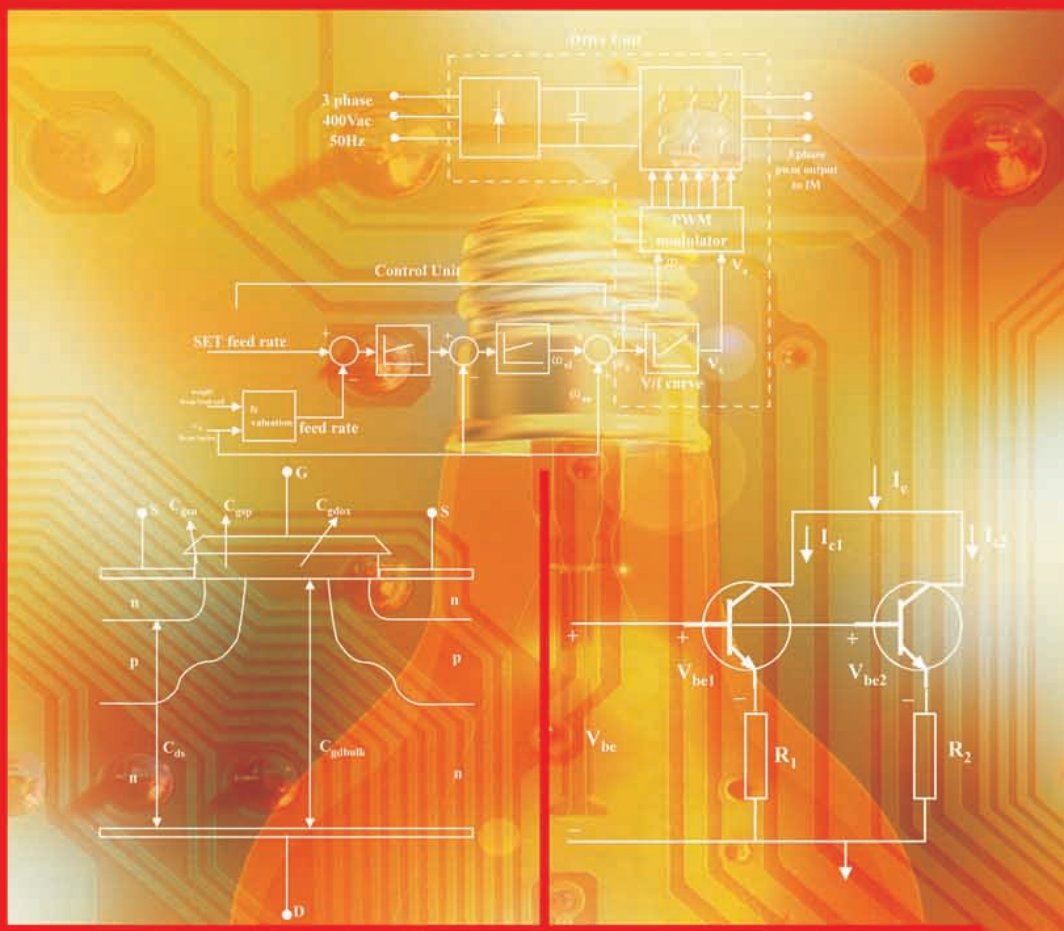
WILEY

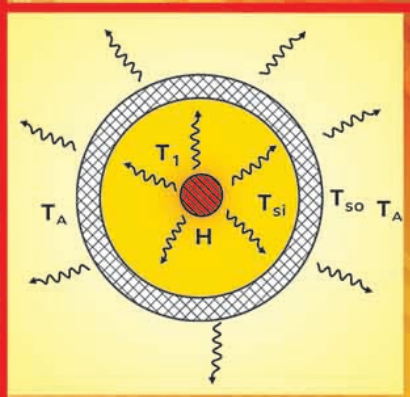
Power Electronics

Essentials & Applications





L UMANAND



VISIT…

Power Electronics

Essentials and Applications

Power Electronics

Essentials and Applications

L. Umanand
Centre for Electronic Design & Technology
Indian Institute of Science
Bangalore 560 012

Power Electronics

Essentials and Applications

Copyright © 2009 by Wiley India Pvt. Ltd., 4435-36/7, Ansari Road, Daryaganj, New Delhi-110002.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or scanning without the written permission of the publisher.

Limits of Liability: While the publisher and the author have used their best efforts in preparing this book, Wiley and the author make no representation or warranties with respect to the accuracy or completeness of the contents of this book, and specifically disclaim any implied warranties of merchantability or fitness for any particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by sales representatives or written sales materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every individual. Neither Wiley India nor the author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Disclaimer: The contents of this book have been checked for accuracy. Since deviations cannot be precluded entirely, Wiley or its author cannot guarantee full agreement. As the book is intended for educational purpose, Wiley or its author shall not be responsible for any errors, omissions or damages arising out of the use of the information contained in the book. This publication is designed to provide accurate and authoritative information with regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services.

Trademarks: All brand names and product names used in this book are trademarks, registered trademarks, or trade names of their respective holders. Wiley is not associated with any product or vendor mentioned in this book.

Other Wiley Editorial Offices:

John Wiley & Sons, Inc. 111 River Street, Hoboken, NJ 07030, USA

Wiley-VCH Verlag GmbH, Pappellaee 3, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada, M9W 1L1

First Edition: 2009

ISBN: 978-81-265-1945-3

ISBN: 978-81-265-8016-3 (ebk)

www.wileyindia.com

Printed at:

Foreword

The book that you have in your hands is the outcome of years of patient but enthusiastic learning and exploration of the field of power electronics by its author. You may feel a bit apprehensive in coming to terms with more than eight hundred pages on this topic; however, a closer look at the table of contents will reveal the comprehensive coverage of the subject, explored from many angles and with various tools. This book, indeed, reflects well the professional trajectory of its author. Having known Dr L. Umanand from his student days and later as a colleague at the Centre for Electronics Design and Technology, I had the pleasure to interact with him during his entire professional career and share his passion for power electronics. I can clearly recognize in this book that same passion that he progressively developed for various aspects of the subject.

He has taken up the challenge to present here a comprehensive synthesis of the field with the required theoretical approach, combined with the pragmatism that comes from experience accumulated through years of direct involvement in design, development and testing of the power electronics systems. This, together with the originality of certain topics like Bond graphs, Design of Magnetism or Design for Reliability, will surely make this book stand out.

The book reflects equally well the deep interest the author has taken in the last few years of the "teaching/learning" process. The learning objectives of each chapter are clearly stated and the material is illustrated with numerous examples. The reader is also challenged throughout the book with numerous questions, problems, laboratory assignments, that, if carried out sincerely, should ensure proper anchoring of the newly acquired knowledge.

In spite of the limitations of any written material, I wish that through this book Dr Umanand will be able to inspire you in sharing his passion and enthusiasm for the subject as he does so well with his students, even if you miss the immediate proximity of his broad smile and great sense of humor.

Andre Pittet
Chief Project Advisor
CEDT, Indian Institute of Science
Bangalore

Preface

The subject of power electronics has been treated in numerous literatures with various viewpoints and styles. At the outset, it may appear as though one more book is being introduced into the market, which may at most be a marginal update on the existing literature with the same topics that are addressed in other books. Likewise, I would like to reiterate that the topics and concepts covered in this book are not new as one would expect, considering the field of power electronics is rather mature. Notwithstanding the maturity of the field, I must add that the topics in this book are neither old wine in new bottle nor a marginal update. Power electronics that initially started its career in processing power revolved primarily around devices like the vacuum tubes, mercury arc rectifiers and, later on, the thyristors. Power electronics has come a long way since then. Technology has improved by leaps and bounds making the power devices more closely to an ideal switch, and the control of the switches are performed in the discrete domain with complex control structures. Thus, power electronics now spans a very wide knowledge base such as power devices, drives, circuit topologies, magnetics, system modeling, control configurations, digital processing, thermal and reliability aspects. If all these aspects are to be treated in a single book, then the topics will, in general, have a superficial bearing. If the topics are to be treated in depth, then it becomes difficult to accommodate all these topics into a single book. Here in lies the challenge and hence this book.

All the topics discussed in this book have been handed down from my teachers and teachers before them by way of exhortation; and to them, I offer my sincere salutations. In the run up to this book, I must say that I have been indeed very lucky to have been tutored by some of the best minds in the field. Every topic has a flavor of the teaching styles of my mentors. From the confluence of the teaching styles of my teachers, I hope that a style will emerge that I may call my own. If the topics are well treated and addressed, then the credit, I must say, should go to my mentors; and if there are mistakes which I presume there will be some, I am solely responsible and will make efforts to correct them in future editions.

The book has been broadly divided into two types of topics: (a) circuit-oriented aspects and (b) system-oriented aspects. The first seven chapters deal with circuit-oriented aspects of power electronic systems, and Chapters 8–15 deal with system-oriented aspects like controls and reliability. Chapter 1 discusses the power semiconductor switches which are the main building blocks of the power electronic systems. The treatment of the semiconductor switches addresses static characteristics, dynamic characteristics, losses and modeling issues. Chapter 2 discusses the drive circuit requirements of BJTs and MOSFETs. Typical drive circuits and their designs are dealt in this chapter. This chapter also introduces the concepts of the series and shunt snubber circuits for power semiconductor devices that behave as power switches.

Chapters 3–6 deal with the power electronic circuit applications. Chapter 3 is on the topic of AC–DC converters or rectifiers. This chapter discusses single- and three-phase rectification topologies along with a detailed discussion on the popular rectifier–capacitor filter circuit addressing its design aspects. The rectifier–LC filter circuit is also discussed. This is followed by a discussion on controlled rectification. Chapter 4 does not fall into the switched-mode category. This chapter addresses the linear regulators wherein the power semiconductor devices are operated in the linear region. Apart from design, the analysis of the linear regulators by progressively including non-idealities is dealt with in a systematic manner.

Chapter 5 discusses the DC–DC converter application. The primary non-isolated topologies of the DC–DC converters are discussed, followed by isolated converters and other special converters. The discussions in this chapter are based on the steady-state analysis that primarily addresses the component design issues. Chapter 6 handles the DC–AC converters or inverters. The various generic topologies are explained followed by a detailed Fourier series analysis of pulse-width modulation strategies for both single- and three-phase inverters. Chapter 7 is a generic topic on magnetic devices that is needed for all applications in AC–DC, DC–DC, DC–AC conversions and isolated drive circuits. This chapter discusses the general principles of magnetism applied to design of inductors, potential transformers and current transformers.

Chapters 8–15 discuss system-oriented aspects of power electronics. Chapter 8 discusses the different modeling methods and their applications to power electronic circuits and systems in order to obtain a dynamic model. Special emphasis is placed on the circuit averaging method, bond graph method and the space vector methods, including detailed discussions with application examples. Chapter 9 is not strictly related to power electronics, but the topics discussed here are essential for designing controllers. It is a cocktail of many topics such as z-transform basics, digital filters, sampling, analog-to-digital conversion methods and performance specification issues.

Chapter 10 develops a formal and systematic approach towards design of controllers for the power electronic systems, wherein the dynamic models are obtained from the concepts of Chapter 8. Both the classical methods of controller design and the state space methods of controller and estimator design are addressed with more emphasis on digital controller and estimator design. Chapter 11 is the extension of the state space controller and estimator design with focus on optimality and robustness issues.

Chapter 12 discusses an important implementation aspect which is discrete computation methods. As most of the control implementations of the power electronic systems are in the digital domain, a detailed discussion on the numeric formats and practical arithmetic algorithms are addressed. This chapter also discusses the implementation of important components like the PI controller and pulse-width modulators within a digital processor. Chapter 13 discusses yet another important and practical aspect of power electronic system which is the thermal aspect. This discusses the heat transfer mechanisms for conducting the heat away from the junction of the power semiconductor devices to the ambient. The selection of heat sinks for the power electronic applications is addressed in this chapter.

Chapters 14 and 15 are devoted to reliability aspects. Chapter 14 discusses in detail the concepts of modeling systems from the reliability point of view. Chapter 15 presents the methods for predicting the reliability of a system. A formal and systematic method to predict the reliability of circuits by part stress co-variate approach is presented by integrating the functional and life aspects of the circuit specifications at the design stage itself. A MATLAB based toolbox called reliability for electronic circuits (REC) is included in the accompanying CD. This toolbox is developed based on the concepts presented in Chapter 15.

As is evident from the topics and the chapters, the vast area of power electronics cannot be handled in a single semester course of the engineering curriculum. In order to handle the topics for semester duration, the chapters may be re-organized in the following manner that is only suggestive and also not exhaustive.

- Parts of Chapters 1, 3, 4 and 7 are a possible combination for a course on linear power supplies.
- Parts of Chapters 1–3, 5 and 7 are a possible combination for a course on switched-mode DC–DC power supplies.
- Chapters 5 and 7 can be suitable for a course on DC–DC converters.
- Parts of Chapters 1–3, 6 and 7 are a possible combination for a course on inverters.
- Chapters 6 and 7 can be suitable for a course on inverters along with pulse-width modulation strategies.

- Chapters 5, 8–10 can be useful for an advanced-level course on dynamics of DC–DC converters.
- Chapters 8–11 by themselves can be suitable for an advanced-level course on control of power electronic systems.
- Chapters 5, 7, 12 and 13 can be a possible combination for a course on implementation aspects of DC–DC converters.
- Chapters 6, 7, 12 and 13 can be a possible combination for a course on implementation aspects of inverters.
- Chapters 14 and 15 by themselves can be useful for a course on reliability of power electronic systems.
- Parts of Chapters 5, 7, 13–15 can be suitable for a course on reliability design for DC–DC converters.
- Parts of Chapters 6, 7, 13–15 can be suitable for a course on reliability design for inverters.

I must acknowledge that though it appears that I am the sole author of this book, the material in the book has evolved over many years with constant interaction with my teachers, colleagues and students who have directly and indirectly contributed to the knowledge base of the book; my sincere salutations to all of them.

For the past year and a half, I have not given sufficient time to my family and the matters of the home due to the long and late hours spent in writing this book. My wife and son have been very patient and supportive in this aspect, awaiting the time for the completion of the book. I sincerely acknowledge my gratitude for their continuous support.

L. Umanand

Contents

Foreword v

Preface vii

1 Power Semiconductor Switches 1

1.1 Ideal Switch 2
1.2 Diodes 2
Static Characteristics 6
Dynamic Characteristics 7
Diode Classifications 8
Diode Parameters 8
1.3 Surge Current 9
1.4 Thermal Viewpoint 9
Diode Model 10
1.5 Bipolar Power Transistors 11
Static Characteristics 13
Dynamic Characteristics 14
Calculation of Power Dissipation in a Transistor 17
1.6 Safe Operating Area (SOAR) 19
1.7 Paralleling of Transistors 20
1.8 Darlington Connection 21
1.9 Power MOSFETs 22
Static Characteristics 25
Dynamic Characteristics 25
Paralleling of MOSFETS 29
Handling Precautions for MOSFET 29
Remarks on BJT and MOSFET 29
Power Dissipated in a Power MOSFET 30
Model for Power MOSFET 30
1.10 Insulated Gate Bipolar Transistor (IGBT) 30
Static Characteristics 34
Dynamic Characteristics 35
Remarks on IGBT 37
1.11 Thyristor Family 37
Silicon-Controlled Rectifier (SCR) 38
DIACs and TRIACs 42
Gate Turn-OFF Thyristor (GTO) 44
MOS-Controlled Thyristor (MCT) 46
Concluding Remarks 46
Laboratory Exercises 47
Fill in the Blanks 51

Descriptive Questions 52
Problems 53
Answers 53

2 Drive Circuits 55

2.1 Transistor Drive Circuits 55
Turn-ON Behavior 56
Turn-OFF Behavior 56
Characteristics and Classification of the Drive Circuits 58
BJT Drive Circuit-1 59
BJT Drive Circuit-2 59
BJT Drive Circuit-3 60
BJT Drive Circuit-4 61
BJT Drive Circuit-5 61
BJT Drive Circuit-6 62
BJT Drive Circuit-7 64
BJT Drive Circuit-8 65
BJT Drive Circuit-9 65
BJT Drive Circuit-10 66
BJT Drive Circuit-11 67
BJT Drive Circuit-12 67
BJT Drive Circuit-13 69
2.2 MOSFET Drive Circuits 70
MOSFET Drive Circuit-1 71
MOSFET Drive Circuit-2 72
MOSFET Drive Circuit-3 73
MOSFET Drive Circuit-4 73
MOSFET Drive Circuit-5 74
MOSFET Drive Circuit-6 75
MOSFET Drive Circuit-7 75
MOSFET Drive Circuit-8 76
MOSFET Drive Circuit-9 77
MOSFET Drive Circuit-10 78
MOSFET Drive Circuit-11 79
2.3 Snubber Circuits 80
Turn-OFF Snubber or Shunt Snubber 80
Turn-ON Snubber or Series Snubber 83
Concluding Remarks 85
Laboratory Exercises 86
Fill in the Blanks 91
Descriptive Questions 92
Problems 93
Answers 94

3 Rectifiers 95

3.1 Uncontrolled Rectifiers 96
3.2 Rectifier Circuits 96
Single-Phase Circuits 96
Three-Phase Circuits 103
3.3 Capacitor Input Filter 109
Design of Capacitor Input Filter Rectifier 110
Turn-ON Currents and Surge Limiting 117
3.4 Power Factor 121
3.5 Rectifier–LC Filter 124
Output Ripple 127
Turn-ON Current 128
Design Summary 128
3.6 Controlled Rectifiers 128
Single-Phase Power Circuits 129
Three-Phase-Controlled Rectifier Circuits 134
Concluding Remarks 142
Laboratory Exercises 142
Fill in the Blanks 145
Descriptive Questions 146
Problems 147
Answers 148

4 DC–DC Linear Regulators 149

4.1 Generic Linear Regulator 149
Operating Principle 150
4.2 Linear Regulator Topologies 151
Shunt Regulator 151
Modified Shunt Regulator 154
Series Regulator 154
Negative Voltage Regulator 156
IC Linear Regulators 158
Practical Tips 160
Four-Wire Connection 162
Protection 163
Current Regulation 165
4.3 Parameters of the Linear Regulator 166
4.4 Analysis of a Series Regulator 166
First-Level Modeling 166
Effect of Current through R_b 170
Effect of Current through R_1 171
Regulator Parameters 172
Range of Load Resistance R_L 173

	Second-Level Modeling	174
	Influence of Temperature	179
4.5	Analysis of a Current Regulator	180
	Line Regulation	182
	Range of R_L	182
	Second-Level Modeling	183
	Concluding Remarks	184
	Laboratory Exercises	184
	Fill in the Blanks	188
	Descriptive Questions	188
	Problems	189
	Answers	192

5 DC–DC Switched-Mode Converters 193

5.1	Single Pole Double Throw Switch	193
5.2	The Chopper	194
5.3	DC Steady-State Principles	200
5.4	The Three Basic Converters	201
	Step-Down or Buck Converter	201
	Step-Up or Boost Converter	206
	Step-Up/Down or Buck–Boost Converter	213
5.5	Selection of Power Switches	219
	Device Selection	219
	Electrical Stress Rating	221
	Thermal Stress Rating	221
5.6	Continuous and Discontinuous Conduction Modes	221
5.7	Isolated Converters	225
	Forward Converter	226
	Semiconductor Switches	228
	Waveforms	229
	Governing Equations	230
	Forward Converter with Demagnetizing Winding	233
	Dual Switch Forward Converter	236
	Push–Pull Converter	238
	Half-Bridge Converter	243
	Full-Bridge Converter	248
	Flyback Converter	253
5.8	Practical Issues	258
5.9	Other Topologies	258
	Cuk Converter	259
	High-Boost Converter	263
	Soft Switching in Converters	265
	Concluding Remarks	275
	Laboratory Exercises	276
	Fill in the Blanks	281

Descriptive Questions 283
Problems 286
Answers 288

6 DC–AC Switched-Mode Converters 289

6.1 Inverter Topologies 290
Single-Phase Topologies 290
Three-Phase Topologies 296
6.2 Self-Driven Inverters 297
Saturable-Core Inverter 298
Saturable Base Drive Inverter 302
6.3 Driven Inverter 304
Push–Pull Configuration 304
Half-Bridge Configuration 306
Full-Bridge Configuration 308
6.4 Quasi-Square Wave Inverter 309
6.5 Three-Phase Inverter Topologies 312
Six-Step Inverter 314
Current-Controlled Inverter 315
6.6 Pulse-Width Modulation 318
Quasi-Square Modulation 321
End Pulse Modulation 324
Center Pulse Modulation 327
Sinusoidal Pulse-Width Modulation 329
PWM through Phase Modulation 337
Space-Vector Modulation 339
6.7 Pulse–Width Modulation with Harmonic Elimination 350
Harmonic Elimination with Multiple Notches 351
Determination of Notch Angles 352
Staircase PWM 355
Summary of Staircase PWM Design 357
Concluding Remarks 358
Laboratory Exercises 359
Fill in the Blanks 363
Descriptive Questions 365
Problems 367
Answers 368

7 Design of Magnetics 369

7.1 Magnetic Concepts 369
Dissipative Component 371
Flux and Flux Density 372
Volt-Second Balance 374

7.2	Inductor	375
	Inductor Value	375
	Energy Storage	375
	Area Product	377
	Window Area	378
	Cross-Section Area	378
	Design of Inductors	379
	Multiple-Winding Inductors	382
7.3	Potential Transformer	383
	Operating Principle	384
	Turns Ratio	385
	Leakage Flux Linkage	386
	Equivalent Circuit	388
	Area Product	391
	Transformer Design	394
7.4	Current Transformer	398
	Design of Current Transformers	401
7.5	General Notes on Magnetics	403
	Window Utilization Factor	403
	Core Shapes	404
	Polarity and Dot Convention	406
	Magnetic Losses	407
	Skin Effect	412
	Proximity Effect	413
	Concluding Remarks	415
	Laboratory Exercises	416
	Fill in the Blanks	418
	Descriptive Questions	419
	Problems	420
	Answers	421

8	**Modeling of Systems**	**423**
8.1	Input–Output Relations	425
8.2	Differential Equations and Linearization	426
8.3	State Space Representation	428
8.4	Transfer Function Representation	431
	Modeling an Armature-Controlled DC Motor	434
	Poles and Zeros	436
8.5	Block Diagrams	436
	Mason's Rule	440
8.6	Lagrange Method	442
8.7	Circuit Averaging	447
8.8	Bond Graphs	453
	Standard Elements	454
	One-Ports	454

Two-Ports 455
Multi-Ports (Junctions) 457
Rules for the Selection of Causality 457
Steps in Obtaining the System Model 457
Bond Graph Construction 458
Causality Assignment 462
State Equation Extraction 463
Modeling Switched Power Systems 467
8.9 Space-Vector Modeling 473
Space Vectors 475
Representation of Space Vectors in Orthogonal Co-ordinates 476
Space-Vector Transformations 476
Modeling of Induction Motor 479
State Space Representation of the d–q Model of the Induction Motor 483
Concluding Remarks 491
Tutorial Exercises 492
Fill in the Blanks 494
Descriptive Questions 496
Problems 497
Answers 498

9 Control System Essentials 499

9.1 Representation of System in Digital Domain 501
9.2 The z-Transform 502
Physical Significance of z^{-1} 506
9.3 Digital Filter 508
Few Common Digital Filters 512
9.4 Mapping between s-Plane and z-Plane 515
$s = 0 + j0 \rightarrow z = 1$ 515
$s = 0 + j\pi/T \rightarrow z = -1$ 516
$s = 0 - j\pi/T \rightarrow z = -1$ 516
$s = -\sigma + j0 \rightarrow z$ = Real Axis between 0 and 1 517
$s = \sigma + j0 \rightarrow z$ = Real Axis between 1 and Infinity 517
9.5 Effect of Sampling 518
9.6 Continuous to Discrete Domain Conversion 521
Pole–Zero Map 522
Numerical Integration Methods 523
Hold Equivalence 530
Discretizing the State Equation 533
9.7 Control System Basics 535
Performance Specifications 535
Stability 535
Steady-State Accuracy 536
Speed of Response 538
Measures for Performance Parameters 538

9.8 Control Principles 540
9.9 State Space Method 542
Concluding Remarks 543
Tutorial Exercises 544
Fill in the Blanks 545
Descriptive Questions 546
Problems 547
Answers 548

10 Digital Controller Design 549

10.1 Controller Design Techniques 550
10.2 Bode Diagram Method 551
Bode Diagrams 552
Controller Design with Bode Diagram 556
10.3 PID Controller 562
PID Controller Design 567
Integrator Windup in PID Controller 568
Practical PID Controllers 570
10.4 Root Locus Method 573
Root Locus Plot 574
Controller Design 575
10.5 State Space Method 582
Controllability 582
Observability 583
10.6 Full-State Feedback 584
Regulator 585
Tracker 585
10.7 Regulator Design by Pole Placement 587
10.8 Estimator Design 591
Estimator Error Dynamics 592
Separation Principle 593
Current Estimator 594
Reduced Order Estimator 596
10.9 Tracker: Controller Design 599
10.10 Controlling Voltage 600
Operation 605
10.11 Controlling Current 607
Current-Controlled Converter 607
Unity Power Factor Converter 610
Front-End Converter 613
10.12 Control of Induction Motor 619
Scalar Control 619
Vector Control 622
10.13 Output Feedback 631

10.14 Induction Motor Control with Output Feedback 633
Plant Dynamics 636
Anti-Alias Filter Dynamics and Augmentation 636
Controller Dynamics and Augmentation 638
Concluding Remarks 641
Tutorial Exercises 642
Fill in the Blanks 643
Descriptive Questions 644
Problems 646
Answers 647

11 Optimal and Robust Controller Design 649

11.1 Least Squares Principle 650
11.2 Quadratic Forms 651
11.3 Minimum Energy Principle 653
11.4 Least Squares Solution 655
11.5 Weighted Least Squares 658
11.6 Recursive Least Squares 659
LMS Algorithm 660
Kalman Filter 664
11.7 Optimal Control: Linear Quadratic 668
Linear Quadratic Regulator 670
Linear Quadratic Gaussian 672
Linear Quadratic Tracker: Output Feedback 672
11.8 Induction Motor Example 674
11.9 Robust Controller Design 679
Frequency Domain Performance Specifications 681
Modeling Uncertainties 681
Controller Design Steps 682
Speed Controller Synthesis 685
Concluding Remarks 688
Tutorial Exercises 688
Fill in the Blanks 691
Descriptive Questions 692
Problems 693
Answers 694

12 Discrete Computation Essentials 695

12.1 Numeric Formats 695
Fixed-Point Numeric Format 696
Floating-Point Numeric Format 699
General Notes on Number System 702

12.2 Tracking the Base Point in the Fixed-Point System 703
Addition of Numbers 703
Subtraction of Numbers 703
Multiplication of Numbers 703
12.3 Normalization and Scaling 704
Multiplication Algorithm 707
12.4 Arithmetic Algorithms 709
Reciprocal 709
Square Root 711
Reciprocal of Square Root 713
Hypotenuse 713
Polynomial 714
Sine and Cosine 714
Exponential 715
Logarithm 716
Implementation Examples 717
PI Controller 717
Sine and Cosine 719
Pulse-Width Modulation (PWM) 720
Space Vector PWM 724
Over-Modulation 727
Concluding Remarks 729
Tutorial Exercises 730
Fill in the Blanks 732
Descriptive Questions 732
Problems 733
Answers 734

13 Thermal Aspects 735

13.1 Modeling Heat Transfer 736
13.2 Heat Transfer by Conduction 738
13.3 Heat Transfer by Convection 742
Determination of Nusselt Number 743
Free Convection 744
Forced Convection 747
13.4 Heat Transfer by Radiation 753
Reflectance, Absorptance and Transmittance 755
13.5 Heat Transfer by Mass Transport 756
13.6 General Notes on Heat Transfer 758
Mixed Heat Transfer Mechanisms 758
Thermal Resistance of Varying Cross-Sections 759
Mounting of Power Devices on Heat Sink 760
Transient Model for Heat Transfer 761
Best Method of Heat Transfer 762

Concluding Remarks 763
Tutorial Exercises 763
Fill in the Blanks 765
Descriptive Questions 766
Problems 767
Answers 768

14 Reliability Modeling 769

14.1 Reliability Concepts 771
Failure Rate 771
Failure-Time Probability Density Function 773
Mean Time Between Failure (MTBF), Mean Time to Failure (MTTF) 776
14.2 Modeling for Reliability 777
Series System 777
Parallel System 778
Mesh System 786
14.3 Redundant Systems 788
Full-On Redundancy: Multiple-Survivor Sub-System 789
Switching in Redundant Systems 791
Mean Time to Failure 795
Standby Redundancy 797
Standby Redundancy with Imperfect Switching 799
14.4 Redundant Systems with Non-Identical Components 799
14.5 Design for Reliability 804
14.6 Reliability Apportioning/Allocation 813
Method 1: Equal Apportionment Technique 814
Method 2: AGREE Apportionment Technique 814
Method 3: ARINC Apportionment Technique 815
14.7 Inter-Connection and Wiring 816
14.8 General Notes on Designing for Reliability 822
Connectors 822
14.9 Part Derating 828
Concluding Remarks 828
Tutorial Exercises 829
Fill in the Blanks 830
Descriptive Questions 831
Problems 832
Answers 833

15 Reliability Prediction 835

15.1 Similar Equipment Technique 836
15.2 Similar Complexity Technique 836
15.3 Prediction by Function Technique 836

15.4 Parts Count Technique 837
Part Count with Non-Series Model 839
15.5 Part Stress Technique 839
15.6 Environments 842
15.7 Reliability Prediction Process 842
15.8 Component Naming Convention 843
15.9 Component Object Notation 844
15.10 Example Reliability Prediction Process 849
Steps to Design a Circuit 849
15.11 Physics of Failure (PoF) 852
Concluding Remarks 854
Tutorial Exercises 854
Fill in the Blanks 855
Descriptive Questions 855
Problems 856
Answers 857

Appendix I 859
Appendix II 861
Appendix III 863
Appendix IV 865
Appendix V 866
Appendix VI 867

Bibliography 909
Index 911

CHAPTER 1

Power Semiconductor Switches

Learning Objectives

After reading this chapter, you will be able to:

- understand the requirements of an ideal switch and the characteristics of important power semiconductor switches.
- estimate the conduction and switching power losses in various power semiconductor switches.
- model and simulate the power semiconductor switches.

Power electronic applications deal with the flow of power. However, majority of the applications in power electronics are based on switching the power flow in order to improve efficiency. As a consequence, one has to use power semiconductor devices that act as power switches. This chapter discusses various power semiconductor devices with an emphasis on using them as power electronic switches. There are many semiconductor devices that can be used as power switches. Few of the common power semiconductor switches are diodes, bipolar junction transistors (BJTs), metal oxide semiconductor field effect transistor (MOSFETs), insulated gate bipolar transistors (IGBTs), thyristors, gate turn-OFF thyristors (GTOs) and metal oxide semiconductor controlled thyristor switches (MCT). Power MOSFETs and IGBTs have a large role to play in modern power control equipments and hence their study is emphasized.

The power semiconductor switches may be studied from various viewpoints:

1. physics viewpoint;
2. circuit viewpoint;
3. protection viewpoint;
4. drive viewpoint;
5. modeling viewpoint;
6. packaging viewpoint.

The **physics** viewpoint explains the operation and the functional features of the device. The **circuit** viewpoint deals with the static and dynamic characteristics. For reliable operation of the power switches, one must ensure that the electrical and thermal stresses within the device are well below the stated ratings of the device. These switches handle large currents and dissipate a considerable amount of heat; consequently, the thermal aspects need detailed attention to ensure that the switch operates within the permissible junction temperatures. The **protection** viewpoint focuses on the electrical and thermal stresses within the

device. In this respect, the safe operating limits for the operation of the power semiconductor device are addressed. The **drive** viewpoint emphasizes and clarifies the switching behavior of the power devices that enables one to synthesize reliable and meaningful drive circuits. The **packaging** viewpoint focuses on the device mounting strategies, removal of heat through heat sinks, forced cooling devices and connection issues.

Modeling switches for purposes of simulation are also important aspects that need to be addressed. The mathematical representation of the switches will aid in simulating the power topologies so that the circuit behavior and the waveforms at various parts in the circuit may be determined before the actual hardware implementation. In fact, it is essential to complement the hardware bread-boarding sessions with electronic bread-boarding sessions not only to enhance the understanding of the circuit and its operation, but also to reduce the design cycle time.

1.1 Ideal Switch

This chapter focuses on semiconductor devices that are operated in such a manner that they behave as switches. These semiconductor switches are supposed to emulate the operation of an ideal single pole single throw (SPST) switch. However, it will become clear from the discussions in the sections to follow that none of the semiconductor switches have all the characteristics of an ideal SPST switch. Before discussing the features of the semiconductor switches, it is necessary to have a reference list of features that an ideal SPST switch has in order to aid in a better understanding of the practical semiconductor switch. An ideal SPST switch will have the following features:

1. ON resistance = 0 (or zero forward voltage drop).
2. OFF resistance = infinity (or zero reverse current).
3. When ON-conducts infinite current in both the forward and reverse directions.
4. When OFF-withstands infinite forward and reverse voltages.
5. It can switch instantaneously from OFF to ON and from ON- to OFF-states.
6. Power dissipated in the switch is zero, that is, both the conduction and the switch transition losses are zero.
7. ON-to-OFF and OFF-to-ON transitions of the switch are fully controllable.
8. It requires no power to drive or control the switch.

While none of the power semiconductor switches have these ideal characteristics, efforts are continuously made to improve the performances of the semiconductor switches such that they may tend towards ideal behavior. There are many different types of power semiconductor devices available commercially. However, in this chapter, only a few popular generic power semiconductor switch types will be discussed. The discussion will initially try to provide some basic insights into the physics of operation of the semiconductor junction taking diode as an example. Subsequently the circuit viewpoint is discussed for the diode as well as few other device types like BJTs, MOSFETs, IGBTs and thyristors to aid in the selection of the devices as power switches for power electronic applications.

1.2 Diodes

Diodes are devices that have two terminals, namely, (a) anode and (b) cathode. When the anode is more positive than the cathode, the diode is said to be **forward-biased** and allows a flow of current from the anode to the cathode. The forward-biased state of the diode is called the ON-state or the conducting state

of the diode. On the other hand, when the cathode is more positive than the anode, the diode is said to be **reverse-biased** and does not permit any flow of current. The reverse-biased state of the diode is called the OFF-state or the blocking state of the diode. The switching action of the diode is solely dependent on the anode-to-cathode potential and this is determined by the external circuit. As there is no control on the switching state, a diode is termed as an **uncontrolled** switch.

In a metal (e.g. copper), the valence electrons are completely free and roam from atom to atom. In a semiconductor (e.g. silicon), a covalent bond is imposed between the valence electrons of adjacent atoms. As a result the motion of valence electrons is coordinated with the motion of valence electrons of an adjacent atom. Thus, in a pure semiconductor there are very few free electrons. Evidently, the conduction property of a metal and a semiconductor is very different. If a fraction of a volt is applied across a semiconductor, a small amount of current will flow whereas if applied across a metal, a large current will result.

The conductivity of pure semiconductors can be changed by adding impurities. This process of adding impurities into a pure semiconductor is called **doping**. By adding some impurities in a pure semiconductor, its conductivity can be increased and electron (or hole) flow can be easily controlled, for example, pure silicon plus a pentavalent impurity (e.g. arsenic) increases the mobile free electron charges and is called ***n*-material**, whereas pure silicon plus a trivalent impurity (e.g. boron) increases the mobile hole charges and is called as ***p*-material**.

A *p*–*n* junction diode can be formed by growing a single crystal of semiconductor material and doping with the above impurities in a controlled manner. Figure 1.1 shows a schematic diagram of a *p*–*n* junction. The *p*-side has a high density of holes (i.e. electron vacancies) and the *n*-side has a high density of electrons. These electrons will diffuse from the *n*- to the *p*-side and the holes will diffuse from the *p*- to the *n*-side. This will create a space charge layer on either side of the junction called the depletion region wherein the holes and the electrons combine to form immobile charges as shown in Figure 1.1. This gives rise to an electric field and a potential barrier. The electric field creates a drift current that acts in such a way as to oppose the diffusion current. Equilibrium is reached when the diffusion and the drift currents balance each other.

The depletion region at the junction has only immobile charges. The value of the potential barrier in volts depends on the charge carriers and is governed by the Boltzmann's relation given in Eq. (1.1). The

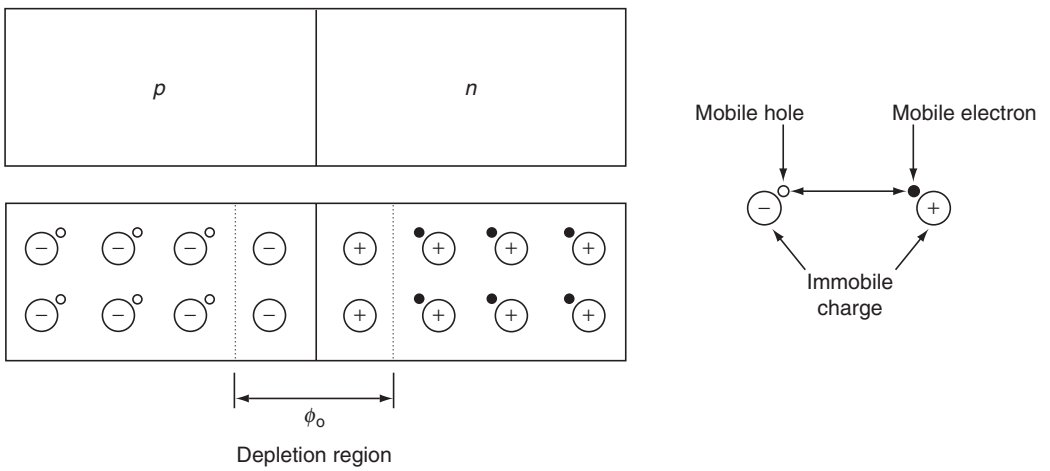


Figure 1.1 | Schematic diagram of *p*–*n* junction.

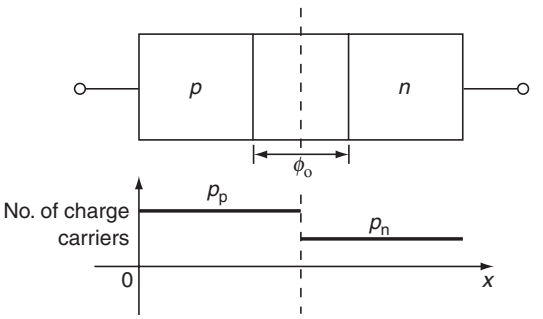


Figure 1.2 | Charge carrier levels in the *p*- and *n*-regions in the absence of external bias.

charge carrier levels at equilibrium for the *p*-region and the *n*-region in the absence of any external bias voltage are depicted in Figure 1.2. The number of holes in the *p*-region is denoted by p_p and the number of holes in the *n*-region is denoted by p_n:

$$p_n = p_p \times e^{-\phi_o/(KT/q)} = p_p e^{-\phi_o/V_T} \tag{1.1}$$

where $V_T = KT/q$ and K is the Boltzmann constant in Joules per degree Kelvin; T is the junction temperature in degree Kelvin; q is the electron charge in coulomb; p_n is the number of holes in the *n* materials; p_p is the number of holes in the *p* material; ϕ_o denotes the barrier potential.

Note that Eq. (1.1) is an exponential relationship. V_T is of the order of few tenths of volt. With this basic relationship one can try to understand the operation of a *p–n* junction that is the building block of almost all semiconductor devices.

On applying a forward bias (positive to *p* and negative to the *n*) to the *p–n* junction as indicated in Figure 1.3(a), the barrier potential across the *p–n* junction reduces from the equilibrium value of ϕ_o to $(\phi_o - \phi)$ and the resulting equilibrium hole density in the *n*-material is changed, which is given by Boltzmann law as

$$p_n^* = p_p \times e^{-(\phi_o - \phi)/(KT/q)} = p_n e^{\phi/V_T} \tag{1.2}$$

Equation (1.2) is a fundamental relationship in junction theory. This shows that a small forward bias increases the minority carrier (holes) concentration in the *n*-region exponentially. These excess holes come from the *p*-side and are in turn replenished from the external source. As the holes cross over from the *p*-region to the *n*-region, swift recombination results. The charge density decreases with distance, *x* from the junction as depicted in Figure 1.3(a). For every electron in the *n*-material that combines with a hole, there is another electron entering the *n*-region from the negative side of the external voltage source. It is important to understand that this change of charge density with distance in the semiconductor involves a transport of charge and thereby constitutes a current flow. It takes place predominantly by diffusion.

The hole-current through diffusion is proportional to dp_n^*/dx. Similarly diffusion current due to the electron density difference also exists. The diffusion current in the *p–n* junction is given as

$$I = I_o(e^{qV/KT} - 1) \tag{1.3}$$

where I is the diode current; I_o the saturation current; q the electronic charge in coulombs; K the Boltzmann's constant; T the junction temperature in degree Kelvin.

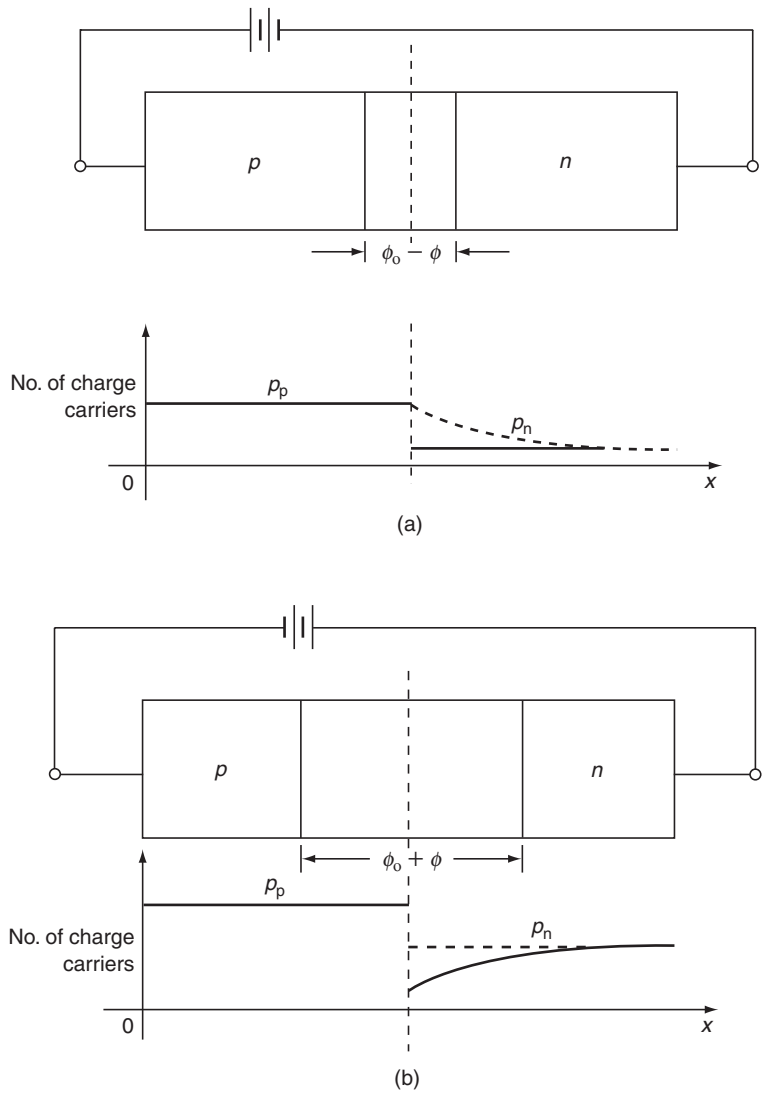


Figure 1.3 | Charge densities near a biased *p*–*n* junction.

In the conventional conductor, the current and the voltage are related by the well-known Ohm's law. However, it is evident from the expression given in Eq. 1.3, that in the case of a *p*–*n* junction, the current is exponentially related to the voltage as a consequence of the Boltzmann relation. On applying a reverse bias (positive to *n* and negative to *p*) to the *p*–*n* junction as indicated in Figure 1.3(b), the barrier potential across the *p*–*n* junction increases to $(\phi_o + \phi)$ and the resulting equilibrium hole density in the *n*-material is given by the Boltzmann law as

$$p'_n = p_p e^{-(\phi_o + \phi)/(KT/q)} = p_n e^{-\phi/V_T} \tag{1.4}$$

Under reverse-biased condition, p'_n is less than the equilibrium density of holes in the *n*-region. This implies that the holes diffuse from the *n*-region into *p*-region. Thus, an extremely small current called the reverse saturation current results and is denoted as I_0. A very large negative voltage disrupts all covalent bonds and results in a large current. Now it behaves like a metallic conductor and the diode is said to have attained reverse breakdown.

Static Characteristics

The symbol of a diode is shown in Figure 1.4(a). Lead A that is connected to the *p*-region of the *p*–*n* junction is called the anode and lead K that is connected to the *n*-region of the *p*–*n* junction is called the cathode. The *v*–*i* characteristics of an ideal diode are depicted in Figure 1.4(b). From Figure 1.4(b), it can be observed that the voltage across the diode when forward-biased is zero. In forward-biased condition it allows flow of currents from the anode to the cathode. When the diode is reverse-biased, then the diode blocks any flow of current. However, the diode can withstand the applied reverse voltage as shown in Figure 1.4(b).

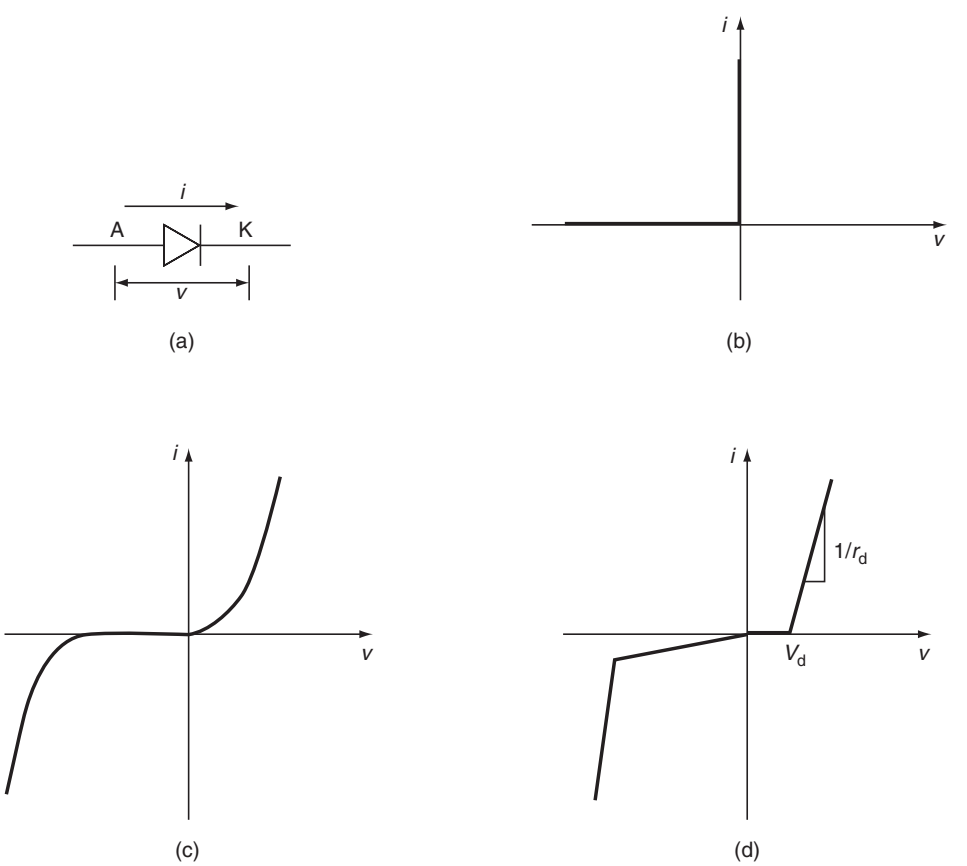


Figure 1.4 (a) Symbol of diode; (b) ideal *v*–*i* characteristics; (c) actual characteristic; (d) piece-wise linear characteristic.

The practical v–i characteristic of the diode is shown in Figure 1.4(c). It shows that during the forward-bias condition, the diode has a finite resistance called the forward dynamic resistance (r_d). For quick engineering calculations, it is normal to simplify the characteristics by the piece-wise linear characteristic, as shown in Figure 1.4(d).

Dynamic Characteristics

Turn-OFF of Diode

A forward-biased diode has a charge distribution as shown in Figure 1.3. If a reverse bias is now applied to the diode, the charges will have to re-distribute from p_n^* to p_n'. This takes a certain amount of time. The waveform for the current and voltage are shown in Figure 1.5. The test circuit used to obtain the dynamic characteristics is as shown in Figure 1.5(a).

The excess charge stored in the diffusion region has to be removed before the junction can be reverse-biased. As long as there are excess charge carriers in the diffusion region (also called the space-charge region), the junction will be in forward-biased state. The diode voltage will not change from its ON-state value except for the small decrease due to Ohmic drop caused by the reverse current. After the current goes negative and the excess charges are removed at time t_2, the junction becomes reverse-biased and quickly acquires the applied negative voltage value. At time t_2, the junction charge distribution would have reached the equilibrium charge distribution of a non-biased junction like that shown in Figure 1.2. From t_2 to t_3

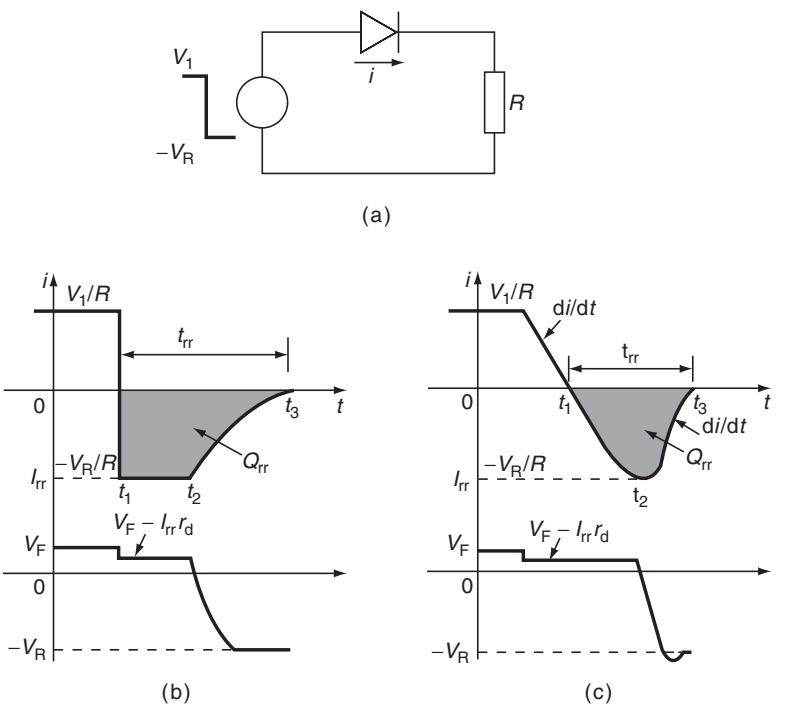


Figure 1.5 Diode turn-OFF: (a) Test circuit for turn-OFF; (b) idealized dynamic characteristic during turn-OFF; (c) practical dynamic characteristic during turn-OFF.

the charge distribution tends towards that of the reverse-biased junction like that shown in Figure 1.3(b). The diode now starts to withstand the applied reverse voltage and the diode current falls quickly to zero. In the case of a practical diode, owing to parasitic inductance, the current in the diode does not immediately reverse like that of the ideal recovery shown in Figure 1.5(b). It reduces at a certain rate decided by the lead inductance as indicated in Figure 1.5(c). The time interval t_{rr} is called the reverse recovery time and is an important parameter in switching applications. The interval between t_1 and t_2 is sometimes called the storage time.

Turn-ON of Diode

If the diode is under reverse bias and it has to be forward-biased, it requires certain time, known as turn-ON time or forward recovery time before all carriers in the whole junction can contribute to the current flow. Here, the charges will have to re-distribute from p_n' to p_n^*. In the case of turn-OFF, the charges have to be removed or re-combined whereas in the case of turn-ON the junction has to acquire charges. This is generally a faster process than removal of the charges. Therefore, turn-ON times will be much faster than turn-OFF times. In practical situations, it turns out that forward recovery time does not constitute a serious problem.

Diode Classifications

The diodes are generally classified based on the turn-OFF times. The reverse recovery time (t_{rr}) is a measure of the speed at which the diode can switch (turn-OFF) and therefore gives an indication on the external switching frequencies that may be used. Based on this, the diodes are classified as follows:

1. Rectifier diodes (t_{rr} in microsecond range).
2. Fast recovery diodes (t_{rr} in 200–500 ns).
3. Ultra fast recovery diodes (t_{rr} ~ 30–200 ns).
4. Schottky diode (metal semiconductor junction $-t_{rr} < 30$ ns).

In high-frequency switching circuits, fast recovery diodes (types 2–4) should be used depending on the frequency range of interest and voltage rating. Type 1 is used for low-frequency applications like mains rectification, as in front-end rectifier for the capacitor input filter.

Diode Parameters

There are various parameters in the datasheet of a diode that need to be understood before selecting a diode for a particular type of application. These parameters must be calculated *a priori* before selecting a diode for the specific application. The important parameters in the datasheet of a diode are

1. Average current, I_{Fav}
2. Root mean square (RMS) current, I_{Frms}
3. Peak current, I_F
4. Surge current, I_{FSM}
5. Breakdown or reverse voltage, V_{RRM}
6. Forward drop, V_F
7. Dynamic resistance, r_d
8. Reverse recovery time, t_{rr}
9. I^2t rating

For sinusoidal currents, I_{Fav}, I_{Frms} and I_F are related and the relationships are well known. It is not so when current in the circuit is non-sinusoidal. In such cases, depending on the current and voltage waveforms of the diodes for the specific circuitry, the corresponding average, rms and peak values must be calculated from the fundamentals.

In high-power circuits, it is important to make sure that the chosen diode ratings are higher than the respective values of currents that will flow in the circuit to ensure that the diode can reliably operate. In detailed datasheets, some of the parameters may be given in the form of nomo-graphs. One may choose the diode based on the parameter nomo-graphs also.

1.3 Surge Current

The rectifier has the ability to sustain overload currents for a brief period of time without damaging the device. Two parameters are specified in datasheets: I_{FSM} and I^2t rating. I_{FSM} is the maximum peak half-cycle non-repetitive current rating. In applications that use the capacitor input rectifier filter, the problem of surge current is severe. When the capacitor is fully discharged, the surge current will be enormous at the time when the rectifier filter is switched ON and the input voltage waveform is at its peak. This problem is most severe in the case of line-operated switched-mode power supplies when the input or the mains voltage is at its worst case maximum value of around 270 V rms. If the expected surge current is higher than what the device can sustain, one has to limit the current using some limiting device such as a series resistor. The amp-squared second (I^2t) rating is normally associated with protective fuse co-ordination and the rating is usually given for 10 ms. For proper functioning, the fuse I^2t rating has to be less than the diode I^2t rating.

1.4 Thermal Viewpoint

An ideal diode does not have any power dissipation. However, a practical diode will have a power loss. It is essential to quantify the amount of power loss in the diodes so that one can select a suitable heat sink in addition to budget the power loss in the circuit so that the efficiency can be estimated. There are different components of power loss in a diode, viz. (a) the ON-state loss (P_{on}); (b) the OFF-state loss (P_{off}) and (c) the switching loss ($P_{switching}$). The total power dissipation (P_d) is given as

$$P_d = P_{on} + P_{off} + P_{switching} \tag{1.5}$$

P_{on} is the average power loss when the diode is in the ON-state. One can use the piece-wise linear model [see Figure 1.4(d)] to estimate the power loss. The power loss is estimated as follows:

$$P_{on} = \frac{1}{T}\int_0^T v \times i \times \mathrm{d}t \tag{1.6}$$

From the piece-wise linear model of the diode depicted in Figure 1.4(d), it can be observed that

$$v = V_d + (i \times r_d) \tag{1.7}$$

where V_d is the cut-in voltage or the knee voltage of the diode as shown in Figure 1.4(d). From Eqs. (1.6) and (1.7), the ON-state power loss in the diode is given as

$$P_{on} = (V_d \times I_{Fav}) + (I_{Frms}^2 \times r_d) \tag{1.8}$$

The knowledge of the average and rms values of the current in the circuit can be used to evaluate the ON-state conduction losses. OFF-state losses are due to the flow of the reverse saturation current

in the reverse-biased p–n junction. The reverse saturation current is negligibly small and therefore the i^2R loss due to it is generally not significant. The reverse recovery losses are negligible at low frequencies (e.g., at 50 Hz operation). However, at higher switching frequencies the reverse recovery losses are significant and affect the efficiency of the circuit considerably. In Figure 1.5(c), the shaded area represents the reverse recovery charge Q_{rr}. The loss during the reverse recovery is given as follows:

$$P_{\text{switching}} = E_{\text{switching}} \times f_s \tag{1.9}$$

where $E_{\text{switching}}$ is the energy spent during the reverse recovery process and f_s is the switching frequency.

$$P_{\text{switching}} = \left(\frac{1}{2} Q_{rr} V_R\right) \times f_s \tag{1.10}$$

Q_{rr} can be obtained from the datasheets and can be used to evaluate $P_{\text{switching}}$. If the shaded portion of Figure 1.5(c) is approximated as a triangle, then Q_{rr} can be estimated from the reverse current during turn-OFF I_{rr} as

$$Q_{rr} = \frac{1}{2} t_{rr} I_{rr} \tag{1.11}$$

Substituting Eq. (1.11) in Eq. (1.10), the switching loss can also be expressed as

$$P_{\text{switching}} = \frac{1}{4} I_{rr} V_R t_{rr} f_s \tag{1.12}$$

I_{rr}, V_R and f_s are circuit parameters which are known for the specific circuit. t_{rr} is obtained from the datasheet for the diode used. It is important to note that the switching losses are a function of the switching frequency. The junction temperature of the device should not exceed a certain limit (150°C as specified in the datasheet). The heat generated in the junction should be removed in such a manner that at thermal equilibrium the junction temperature is well below the rated junction temperature. This is made possible by mounting the device on a heat sink. The thermal issues and the selection of heat sink for a specific calculated device power dissipation are discussed in Chapter 13.

Diode Model

A true model of a diode should take care of the steady-state and transient behaviors. One of the commonly used models for simulation is shown in Figure 1.6. Referring to Figure 1.6, i_d–V_d variables are characterized by an exponential relationship in accordance with Eq. (1.3). R_s is the Ohmic resistance of bulk material plus

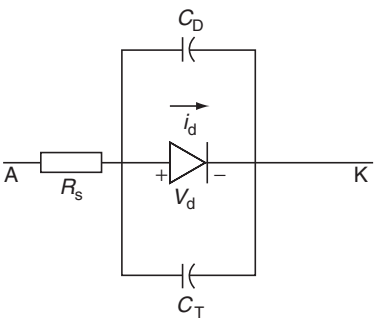


Figure 1.6 | Diode model.

the contact resistance. It is ideally zero; however, in practice it has a finite but small value. When a diode is abruptly switched from the forward to the reverse bias, the excess mobile charge has to be removed. This can be modeled as a diffusion capacitance C_D. For a transition from reverse to forward bias, the mobile charge must build up. This is characterized by a transition capacitance C_T. It should be noted that both C_D and C_T are capacitances whose values are not constant but dependent on the charge in it. Circuit simulator programs also use a similar model.

1.5 Bipolar Power Transistors

The bipolar transistor (bipolar junction transistor, BJT) is a three-terminal device. It has an emitter lead E, a collector lead C and a base lead B. The flow of current from the collector to the emitter is controlled by the current through the base. This base current is a fraction of the current through the collector. Therefore the BJT is called a current-controlled device. In the ON-state, the BJT can allow the flow of current in only one direction (from collector to emitter for NPN transistors; from emitter to collector for PNP transistors). Further, they can support only unidirectional voltages during the OFF-state.

BJT is a dual junction device with two possible configurations, namely, NPN or PNP. Figure 1.7(a) shows a schematic of the vertical cross-sectional structure of the NPN transistor. Figure 1.7(b) shows the circuit symbol of an NPN transistor. Referring to Figure 1.7(a), the transistor consists of a highly doped *n*-type emitter. The base region is a *p*-type doped semiconductor. The bottom layer is the *n*-type collector. The doping of the *n*-type collector is light towards the collector–base junction and higher towards the collector terminal. Generally, the collector is the largest region. Larger the collector region, greater is the amount of voltage that the transistor can withstand during the OFF-state. The arrow in the NPN transistor symbol points out of the emitter. In the case of the PNP transistor, the arrow in the emitter would point inwards as depicted in Figure 1.7(c). In general, the arrow would indicate the conventional direction of the current flow in the device. In the discussions to follow, NPN transistors only will be considered as the concepts of the NPN transistors can be extended to the PNP transistors without loss of generality.

Figure 1.8 depicts the basic operation of an NPN transistor. The base–emitter junction is forward-biased by connecting an external voltage V_E as shown. Electrons are injected from the emitter into the base. They

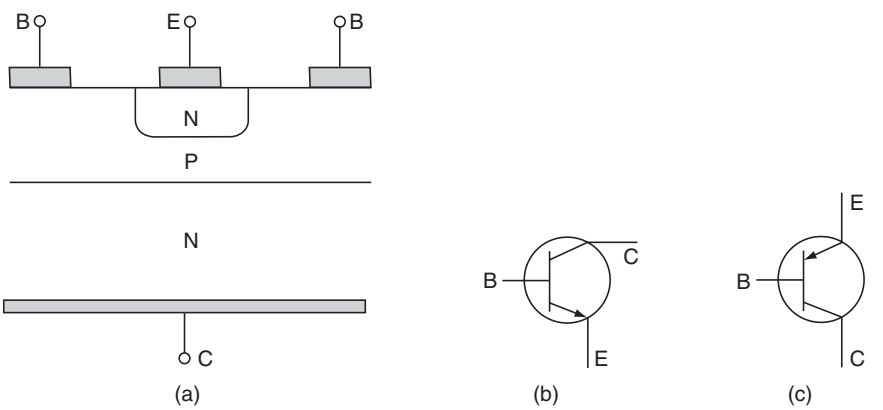


Figure 1.7 | NPN transistor: (a) Structure; (b) symbol (c) Symbol of PNP transistor.

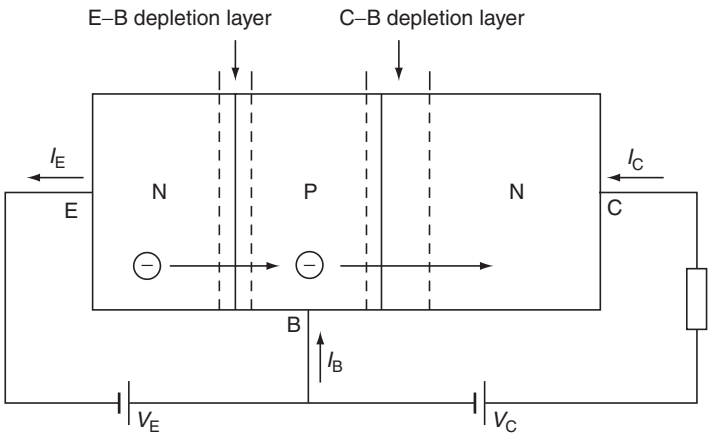


Figure 1.8 | Operation of NPN transistor.

appear at the emitter–base junction. Diffusion takes place in a manner similar to that of the *p–n* junction diode. However, the base is made intentionally small to prevent re-combination of electrons with holes in the *p*-region. This results in a very small I_B. As a result, the vast majority of electrons injected across the emitter–base junction move on to the base–collector junction and further into the *n*-material of the collector. Once the injected electrons appear in this region, the effect of V_C accelerates them towards collector, through the load resistor, through V_C and then back to V_E, thus completing the circuit.

The ratio of the collector current to the emitter current is called the "alpha parameter" of the transistor, that is, $I_C/I_E = \alpha$. This parameter varies between 0.9 and 0.98 depending on the base width. If the base were wide, then all electrons that have crossed the emitter–base junction will re-combine with holes in the base and there will be lesser or no electrons reaching the collector. In this case the emitter–base junction and collector–base junction will act as though they were diodes connected back-to-back. There is no transistor action taking place in such a case. From the discussion above, it should be noted that the transistor is not simply a back-to-back connection of two diodes.

The transistor connected as shown in Figure 1.8 is called the normal mode of operation and α is referred to as α_F or α_N. One can also have an inverted mode of connection wherein the collector acts as the emitter and vice versa. The α under this configuration is denoted as α_R or α_I. One can write the equivalent circuit model of the NPN transistor as shown in Figure 1.9. It is called the Ebers–Moll model. The equations for I_E and I_C for the normal and inverted mode of operations can be written as

$$I_C = \alpha_F I_F - I_{CO}(e^{V_C/V_T} - 1) \tag{1.15}$$

$$I_E = \alpha_R I_R - I_{EO}(e^{V_E/V_T} - 1) \tag{1.16}$$

Referring to the equivalent circuit model of the NPN transistor depicted in Figure 1.9,

$$I_F = I_{EO}(e^{V_E/V_T} - 1);\ \ I_R = I_{CO}(e^{V_C/V_T} - 1);\ \ I_C + I_E + I_B = 0$$

Here I_{EO} is the emitter–base junction saturation current when collector–base junction is zero-biased ($V_c = 0$); I_{CO} is the collector–base junction saturation current when emitter–base junction is zero-biased ($V_E = 0$).

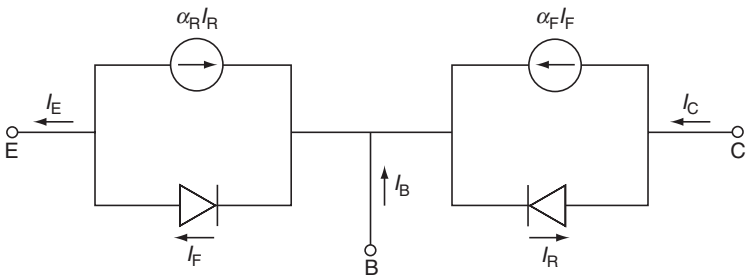


Figure 1.9 | Equivalent circuit of an NPN transistor.

Including the bulk resistance and capacitance (similar to the diode case) results in the model shown in Figure 1.10. C_{DE} and C_{TE} are the emitter–base junction diffusion and transition capacitances, respectively. Similarly, C_{DC} and C_{TC} are the collector–base junction diffusion and transition capacitances, respectively. Circuit simulation programs use a further refined version of the model called as the Gummel–Poon model.

Static Characteristics

BJT is a current-controlled device. Its output characteristics, that is, i_c versus V_{ce} is dependent on the base current i_b. Thus,

$$i_c = f(V_{ce}, i_b) \tag{1.17}$$

where V_{ce} is the collector-to-emitter voltage; i_b is the base current and i_c is the collector current.

Ideally $i_c = \alpha i_e = \beta i_b$ and should be independent of V_{ce}. However, due to "Early effect" (base width changes as V_{ce} changes) the parameter β changes with V_{ce}. Therefore, the i_C versus V_{ce} characteristics change with different values of the base current as depicted in Figure 1.11.

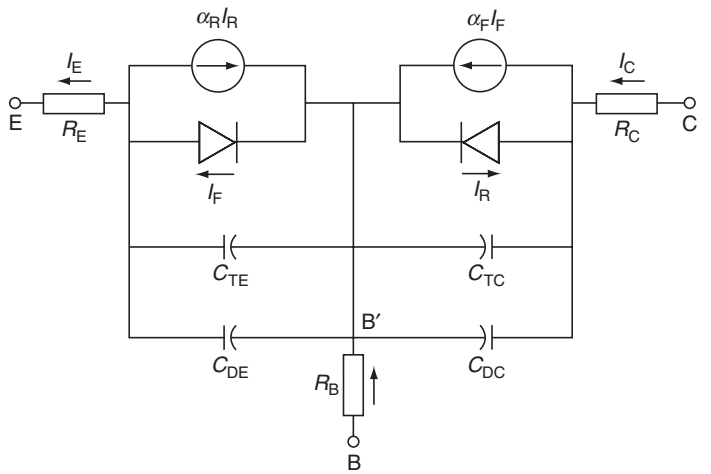


Figure 1.10 | Equivalent circuit of a transistor with bulk resistance and capacitance.

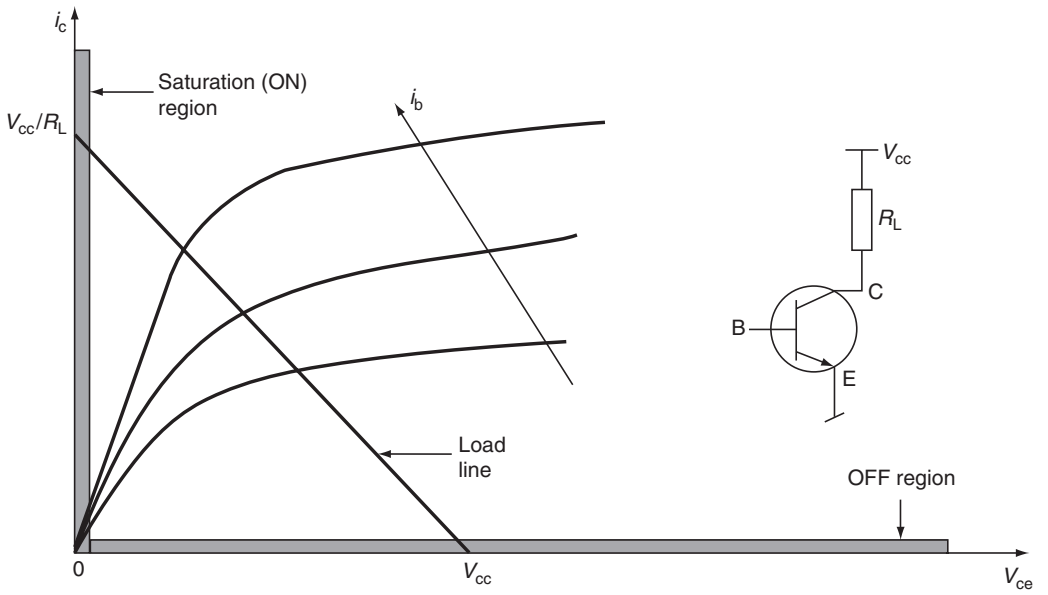


Figure 1.11 | Output characteristics of a transistor.

At very low V_{ce} values, the transistor is said to be in the ON-state. This region is called the saturation region. In this region, the collector current is independent of the base current and depends on the value of V_{ce} and load resistor R_L. A transistor parameter of interest is the ratio of the collector current to the base current, denoted by β or h_{FE}. The parameter h_{FE} is a useful number essential for design and is supplied by the manufacturers. Knowledge of i_C and h_{FE} gives the base current that will be needed to saturate the transistor or maintain the transistor in the ON-state.

At very low collector-current values, the transistor is said to be in the OFF-region as depicted in Figure 1.11. Here the transistor is capable of supporting the applied external voltage. The other region of the output characteristics that is not shaded is called the active region wherein the transistor behaves as a dynamic resistor. For a transistor to be operated as a switch, it is primarily switched between the saturation region or the ON-region and the OFF-region of the characteristics. However, during the transition from the saturation region to the OFF-region and back, the operating point may transit through the active region resulting in a loss called the switching loss.

Dynamic Characteristics

This sub-section describes the switching processes within a transistor under various base drive conditions. For a transistor that is in the ON-state, the following three charges may be said to exist.

1. A charge Q_b located in the base area.
2. A charge Q_{ce} located in the collector region underneath the emitter.
3. A charge Q_{cb} located in the collector region underneath the base contact.

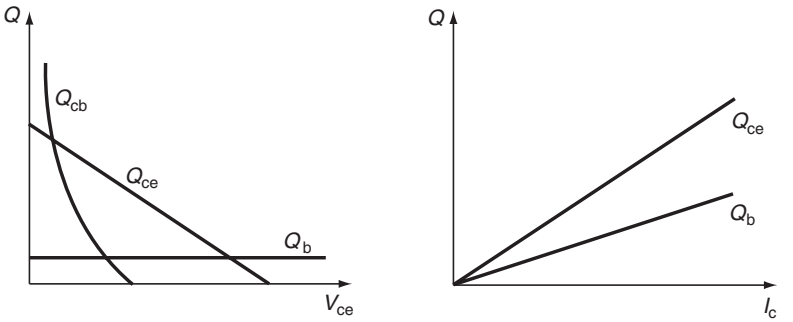


Figure 1.12 | Charge dependency on V_{ce} and I_c.

However, there are no charges present in a transistor that is in the OFF-state. Figure 1.12 illustrates the dependency of these charges with respect to the collector–emitter voltage and the collector current.

Referring to Figure 1.12, it may be observed that the charge Q_b is independent of the collector–emitter voltage drop. However, Q_b increases linearly as collector current increases. The charge Q_{ce} also increases linearly with the collector current but decreases linearly with increases in collector–emitter drop. To achieve lower collector–emitter drop, Q_{ce} must increase. The charge Q_{cb} that is located in the collector underneath the base contact has a significant effect on the collector–emitter voltage as depicted in Figure 1.12. The charge Q_{cb} increases rapidly as V_{ce} drop decreases and the transistor goes towards saturation.

Turn-OFF of Transistor

During turn-OFF, the charges in the transistor must be removed. For a transistor that is in the ON-state, the charges Q_b, Q_{ce} and Q_{cb} are as depicted in Figure 1.13(a). A collector current will flow only when there is a base charge Q_b. The charge Q_{ce} will be located in the collector region underneath the emitter. This results in a low Ohmic collector consequently making the collector–emitter voltage low. An increase in the charge, Q_{ce} results in a decrease in V_{ce}. The charge Q_{cb} will be located in the collector region underneath the base contact. When the base–collector region becomes forward-biased, Q_{ce} rapidly increases and V_{ce} decreases.

The process of turn-OFF of the transistor is illustrated in Figure 1.13. Referring to Figure 1.13(a), when the base voltage is made zero or negative, a negative base current flows as the collector potential is much greater than the base potential. This removes the charge Q_{cb}. Then the charge Q_c will be removed, starting from the area underneath the edges of the emitter as shown in Figure 1.13(b). Q_b also will now start to decrease.

As Q_b starts to decrease, the collector current will gradually be forced towards the center as shown in Figure 1.13(c). As long as Q_b is sufficiently high, the collector circuit will force the collector current to flow. However, V_{ce} will increase now as the current density becomes higher. With the emitter current reducing, the negative base-current flows through the base resistance underneath the emitter (R_b – base-spreading resistance). The charges Q_{ce} and Q_b now become increasingly located beneath the center of the emitter and thus have to be extracted through an increasing resistance. Consequently, the base–emitter terminal voltage becomes more negative. Subsequently, the emitter current is concentrated in the middle of the emitter. This marks the end of a period called the storage period wherein most of the stored charges in the base are removed. The storage period is denoted as storage time t_s and is in general given in the transistor datasheet.

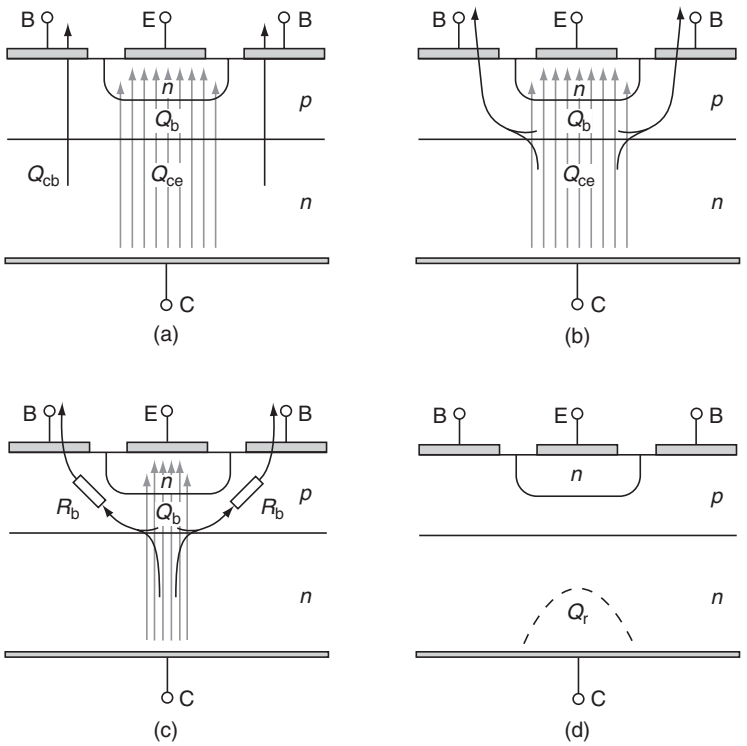


Figure 1.13 | Turn-OFF of the transistor.

Fall time t_f begins as soon as Q_b is so low that the emitter injection of electrons starts to reduce. The emitter current decreases with a speed depending on the rate of decrease in Q_b. $I_e = 0$ when $Q_b = 0$. A trapped rest charge Q_r in the collector must still be removed by way of a collector–base current which appears as a tail current in the turn-OFF waveforms. This rest charge is indicated in Figure 1.13(d).

Turn-ON of Transistor

A transistor that is in the OFF-state has no charges within it. To obtain a low ON-state voltage, a base current is applied such that the collector charge Q_{ce} is built up, which will accordingly reduce the collector resistance and therefore the collector–emitter voltage (see Figure 1.12). To quickly build up the collector charge, the base-current waveform should have a peak at the beginning.

The dynamic characteristics of the transistor are shown in Figure 1.14. Referring to Figure 1.14(a), during the fall time t_f, the voltage across the collector–emitter [shown shaded in Figure 1.14(a)] is decided by the external circuit. If the load is resistive, then the voltage will rise linearly as the collector current falls linearly. If the load were an inductive one, then the voltage during the t_f period would be governed by $L\mathrm{d}i_L/\mathrm{d}t$, where L is the inductance of the load and i_L is current through the load inductance.

Referring to Figure 1.14(b), during the rise time t_r, the current during this time [shown shaded in Figure 1.14(b)] is decided by the external circuit. If the load is resistive, then the current during this time will rise linearly. If the load across the collector–emitter of the transistor is capacitive, the current during the period t_r is governed by $C\mathrm{d}V_{ce}/\mathrm{d}t$, where C is the capacitance across the collector–emitter.

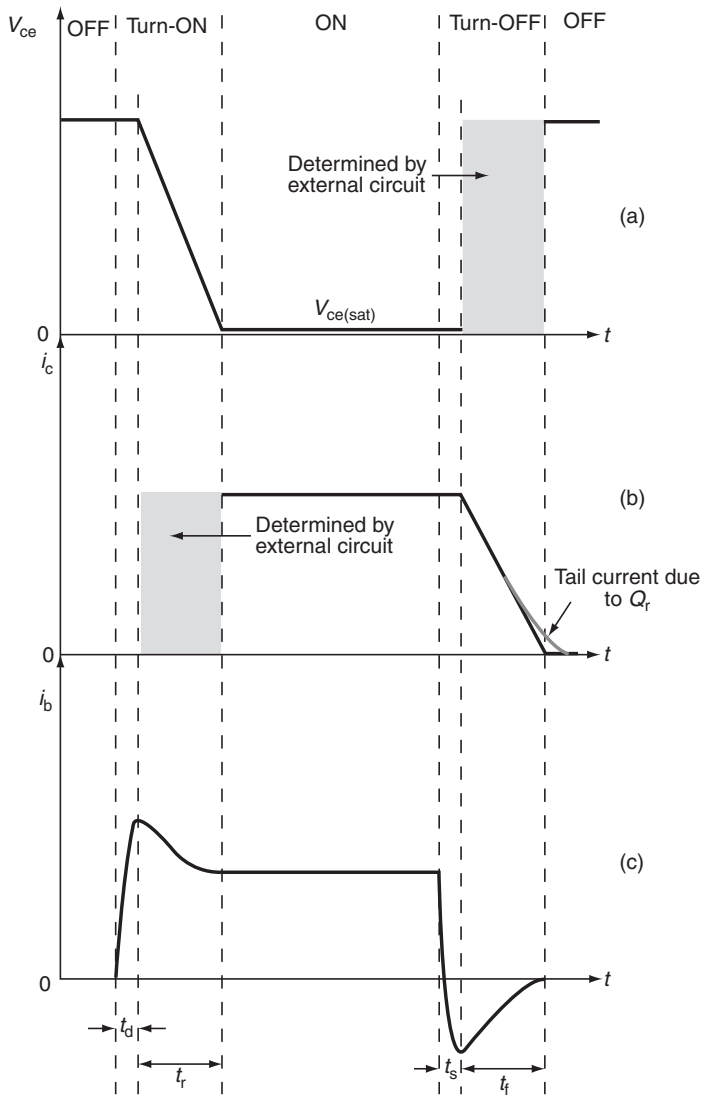


Figure 1.14 Switching characteristics of a transistor: (a) Collector–emitter voltage waveform; (b) collector current; (c) base current.

Calculation of Power Dissipation in a Transistor

The different components of power loss in a transistor are: (a) the ON-state loss (P_{on}); (b) the OFF-state loss (P_{off}) and (c) the switching loss ($P_{switching}$). The total power dissipation (P_d) is given by

$$P_d = P_{on} + P_{off} + P_{switching} \tag{1.18}$$

P_{on} is the average power loss when the transistor is in the ON-state. The power loss is estimated as follows:

$$P_{on} = \frac{1}{T}\int_0^T v \times i \times dt \tag{1.19}$$

where v is the voltage across the collector–emitter of the transistor and i is the collector current through the transistor. During the ON-state, the voltage across the transistor, $v = V_{ce(sat)}$. Therefore Eq. (1.19) becomes

$$P_{on} = V_{ce(sat)} \times \frac{1}{T}\int_0^T i \times dt = V_{ce(sat)} I_{cavg} \tag{1.20}$$

For applications wherein the collector current is a pulsed current with a flat top I_c during the ON-state, the ON-state loss is given by

$$P_{on} = V_{ce(sat)} \times I_c \times D \tag{1.21}$$

where D is the duty cycle given by the ratio of the ON-time to the total switching period.

Like in the case of the diodes, the OFF-state losses are generally negligible. In the case of the switching losses, $P_{switching}$ depends on the nature of the load as is evident from the dynamic characteristics shown in Figure 1.14. A representative example case is discussed wherein the load is resistive. For resistive loads, assuming a linear rise and fall of voltages and currents, the switching losses can be calculated as discussed in the following sub-sections.

Switching Loss from OFF-State to ON-State

Referring to Figure 1.14, when the base drive signal is initiated, there is a time duration (t_d) called the delay time during which time the voltage and the current waveforms do not change much as the charges Q_b, Q_{ce} and Q_{cb} are being set up. This is followed by the time duration t_r. During the time t_r, the voltage across the device falls from V_{cc} to $V_{ce(sat)}$. Without much error, this fall may be assumed to be linear for a resistive load. The collector–emitter voltage during the rise time t_r is

$$v_{ce} = V_{cc}\left(1 - \frac{t}{t_r}\right) \tag{1.22}$$

The collector current for a resistive load will linearly rise from zero to I_c. This is given as

$$i_c = I_c \cdot \left(\frac{t}{t_r}\right) \tag{1.23}$$

The power dissipation during the OFF-state to ON-state transition is given by

$$P_{OFF\text{-}ON} = \int_0^{t_r} V_{cc}\left(1 - \frac{t}{t_r}\right) I_c \left(\frac{t}{t_r}\right) dt = \frac{V_{cc} I_c t_r f_s}{6} \tag{1.24}$$

where f_s is the switching frequency.

Switching Loss from ON-State to OFF-State

Referring to Figure 1.14, when the base drive signal is stopped, there is a time duration t_s called the storage time during which the voltage and the current waveforms do not change much as the charges Q_b, Q_{ce} and Q_{cb} are being removed. This is followed by the time duration t_f. During the time t_f, the current through the device falls from I_c to zero. Without much error, this fall may be assumed to be linear for a resistive load. The collector current during the fall time t_f is given as

$$i_c = I_c\left(1 - \frac{t}{t_f}\right) \tag{1.25}$$

The collector–emitter voltage for a resistive load will linearly rise from $V_{ce(sat)}$ to V_{cc}. This is given as

$$v_{ce} = V_{cc}\left(\frac{t}{t_f}\right) \tag{1.26}$$

The power dissipation during the ON-state to OFF-state transition is given by

$$P_{ON\text{-}OFF} = \int_0^{t_f} I_c\left(1-\frac{t}{t_f}\right)V_{cc}\left(\frac{t}{t_f}\right)dt = \frac{V_{cc}I_c t_f f_s}{6} \tag{1.27}$$

where f_s is the switching frequency. The total switching loss is given as

$$P_{switching} = P_{OFF\text{-}ON} + P_{ON\text{-}OFF} \tag{1.28}$$

Substituting Eqs. (1.24) and (1.27) in Eq. (1.28), one obtains

$$P_{switching} = \frac{V_{cc}I_c f_s(t_r + t_f)}{6} \tag{1.29}$$

It should be observed that the switching losses are proportional to the frequency of switching. The load is generally never resistive. If the load is inductive, the voltage across the transistor during turn-OFF is determined by $L\mathrm{d}i_c/\mathrm{d}t$, where L is the external load inductance. If the load is capacitive, the current through the transistor during turn-ON is determined by $C\mathrm{d}V_{ce}/\mathrm{d}t$, where C is the load capacitance as seen at the collector. The switching losses must be estimated for non-resistive loads also in a manner similar to that discussed for the resistive load. The total power dissipated is the sum of the ON-state loss and the switching loss. The power dissipated as above must be transported away from the junction such that the junction temperature remains at a safe value in equilibrium conditions. A proper heat sink has to be selected. The thermal calculations are discussed in Chapter 13.

1.6 Safe Operating Area (SOAR)

The maximum electric field across the collector–emitter of the transistor must remain below a critical value at all instants for proper functioning of the transistor. The electric field across the transistor is dependent on the collector-current density and the applied collector–emitter voltage. The electric field increases with increasing collector voltage. It also increases with increasing collector-current density. If the collector voltage is lowered then a higher collector-current density is permitted and vice versa. During switching transitions, there are some destructive combinations of the collector-current density and the collector voltage that are likely to occur. The SOARs give information about a given device on the current and voltage handling capabilities.

The collector-current density is dependent on the collector current and the amount of current crowding in the regions of the collector. The amount of current crowding is different for turn-ON (positive base voltage) and turn-OFF (negative base voltage) conditions. Therefore, the allowed combinations of the collector current and collector voltage will differ for turn-ON transition and turn-OFF transition. This information is available in the forward safe operating area (FSOAR) and the reverse safe operating area (RSOAR). A transistor in the ON-state or the OFF-state has the operating point along the Y-axis or the X-axis when either V_{ce} or i_c is zero. However, during switching transitions, both V_{ce} and i_c are non-zero and finite. This means that the operating point will be in the region of the I-quadrant of the V_{ce}–i_c characteristic. In selecting a transistor, the one which operates within both the FSOAR and the RSOAR, for the specific circuit, should be chosen.

With positive voltage applied to the base, the shape of a typical SOAR characteristic is as shown in Figure 1.15(a). The solid line shows the SOAR for DC operation and the dashed lines shows the SOAR for pulsed operation. Operation outside the safe operating area is not allowed. For pulsed operation, the FSOAR increases

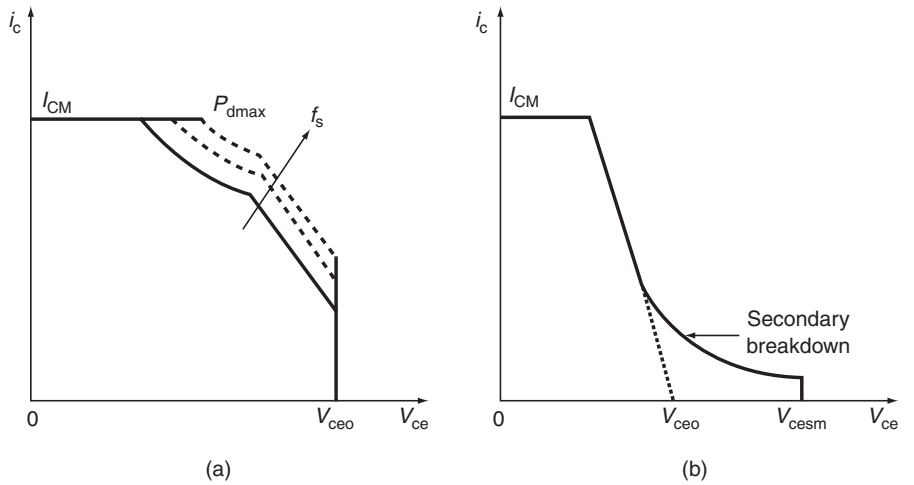


Figure 1.15 | Safe operating area (SOAR): (a) Forward SOAR (FSOAR); (b) reverse SOAR (RSOAR).

and for very small duty cycles, the FSOAR becomes square. The SOAR is designed to indicate the current, power dissipation, voltage and second breakdown limits of the transistor as depicted in Figure 1.15(a). The power dissipation limit is the hyperbolic portion of the limit curve that is given by the maximum possible dissipation allowed for the device. The second breakdown limit is the straight sloped line connecting the power dissipation limit and the voltage limit lines. The second breakdown is generally triggered by combinations of high collector-voltage and high collector-current density. With a positive voltage applied to the base, the region of highest current density is at the edge of the emitter which conducts a substantial proportion of the collector current. During sudden change of currents (especially during switching), a thermal gradient gets generated across the cross-section of the collector–emitter current flow. This thermal gradient will result in the non-uniform spreading of the current in the device. This uneven spreading produces localized hot spots and it can in turn reduce the local resistance, further increasing the non-uniformity. This will result in hot spots in the device that will finally destroy the device.

During turn-ON of the transistor, the high resistance of the collector region is reduced by the introduction of holes from the base and electrons from the emitter. This process is known as conductivity modulation. However, during turn-OFF of the transistor, these extra holes and electrons constitute a stored charge that must be removed from the collector before the voltage across the depletion region can develop. To turn OFF the transistor, a negative voltage is applied to the base and reverse base-current flows. During turn-OFF, it is essential that the device stays within its reverse safe operating area (RSOAR) that is shown in Figure 1.15(b).

1.7 Paralleling of Transistors

In high-power switching applications, transistors can be paralleled to share the load current. The transistor has a negative temperature co-efficient of resistance. As a result, one has to ensure that the current sharing between various parallel transistors is uniform, such that no transistor is over burdened and goes into a thermal runaway. It is very common to add an emitter resistance as shown in Figure 1.16 to equalize the currents. The emitter resistors give a negative feedback in the base–emitter circuit. If I_c increases then V_{be} will decrease, which would result in a decrease of I_b. This consequently will counteract the increase of I_c. The emitter resistor values are governed by the following equations:

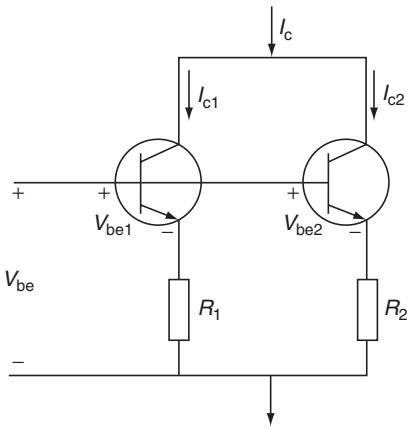


Figure 1.16 | Paralleling of transistors.

$$V_{be} = V_{be1} + RI_{c1} = V_{be2} + RI_{c2} \tag{1.30}$$

where $R_1 = R_2 = R$. From Eq. (1.30), one obtains

$$\Delta V_{be} = V_{be1} - V_{be2} = R\Delta I_c \tag{1.31}$$

Limit ΔV_{be} to around 0.2 V. Let ΔI_c (difference in collector currents between the two transistors) of about 0.5 A be tolerated. This means $R = 0.2/0.5 = 0.4\ \Omega$. It should be noted that R dissipates power. Higher R means more dissipation and higher V_{be} (and V_{cc}) requirements. Paralleling transistors simplify heat transfer problem by spreading the heat produced in a group of transistors over a large surface area heat sink.

1.8 Darlington Connection

High-power, high-voltage transistors generally have a low h_{FE} (β). The h_{FE} of saturated high-voltage transistors can be as low as 2. This means a very high base drive current is required. Therefore, the base drive circuit is no longer a simple low-power circuit. To circumvent this problem, Darlington connection as shown in Figure 1.17 can be used. In this case, the I_b requirement can be cut down to $I_c/(h_{FE}{}^* h_{FE})$. It should

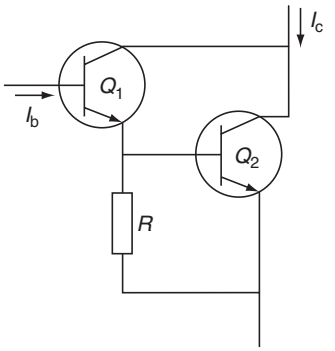


Figure 1.17 | Darlington connections of transistors.

be noted that $V_{ce(sat)}$ of the Darlington device is higher (~1 V to 1.2 V) as opposed to 0.4–0.6 V $V_{ce(sat)}$ of a single transistor. This implies that the power dissipation is higher in Q_2 for a given load.

1.9 Power MOSFETs

The metal oxide semiconductor field effect transistor (MOSFET) is a three-terminal device. It has a source lead S (analogous to the emitter of the transistor), a drain lead D (analogous to the collector of the transistor) and a gate lead G (analogous to the base of the transistor). The flow of current from the drain to the source is controlled by the voltage applied between the gate–source terminals. Thus the MOSFET is a voltage-controlled device. In the ON-state, the MOSFET can allow current in only one direction (from drain to source for *n*-channel MOSFETs and from source to drain for *p*-channel MOSFETs). In the OFF-state, they can support only unidirectional voltages in a manner similar to the BJT.

A structural schematic of an *n*-channel power MOSFET is shown in Figure 1.18(a). It consists of three layers, viz., (a) an *n*-type semiconductor that is connected to the drain D; (b) an *n*-type semiconductor that is connected to the source S and (c) in between the drain and the source is a *p*-type semiconductor that is on the substrate. Generally the substrate is connected to the source within the MOSFET. The gate is connected to a metallic conductor. This is insulated from the bulk of the MOSFET by an insulator that is generally an oxide of some metal such as silicon dioxide. The symbol for the *n*-channel MOSFET is shown in Figure 1.18(b) and the symbol for the *p*-channel MOSFET is shown in Figure 1.18(c). Note that for the *n*-channel MOSFET, the arrow mark shown in the symbol points inwards towards the gate and in the case of a *p*-channel MOSFET, the arrow mark points outwards, that is, away from the gate.

When a positive voltage is applied at the gate as shown in Figure 1.19, an electric field will be generated across the insulating oxide layer. This brings about a polarization of charges within the oxide layer. Consequently, these charges attract electrons from the *p*-material thereby creating an induced *n*-channel that

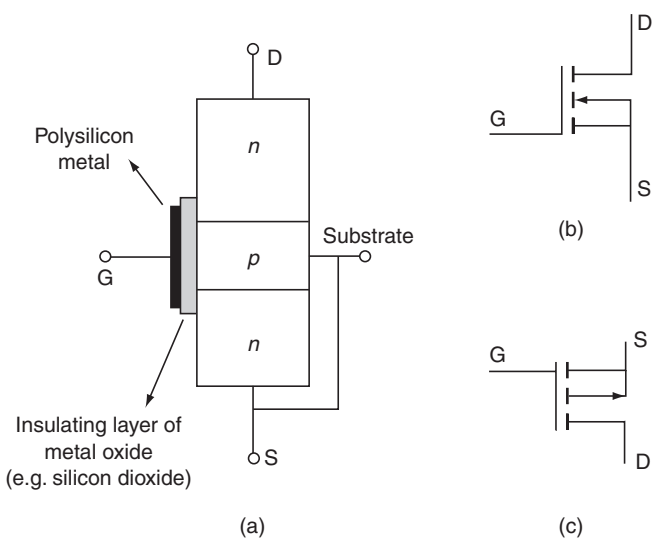


Figure 1.18 (a) Structural schematic of an *n*-channel MOSFET; (b) symbol of an *n*-channel MOSFET; (c) symbol of a *p*-channel MOSFET.

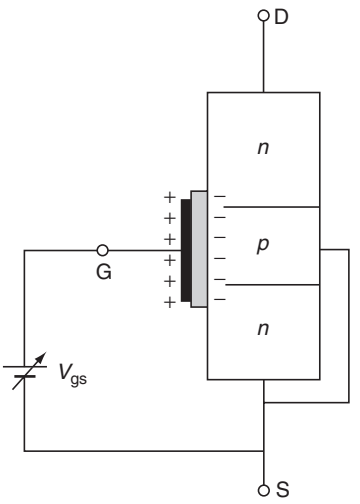


Figure 1.19 | Operation of the MOSFET.

bridges the *n*-type drain and source regions of the MOSFET. If a positive drain voltage is now applied between D and S terminals, the *n*-channel bridge provides a conducting path between the two *n*-regions. This conducting path that bridges the two *n*-type semiconductors in the MOSFET is called the **inversion layer**. As the voltage V_{gs} is increased, the width of the inversion layer increases. The MOSFET is said to have reached full enhancement when the inversion layer width is maximum.

It can be noted that there is a parasitic NPN BJT between the drain and source contacts with *p* substrate serving as the base of the parasitic BJT. To minimize the possibility that this transistor is ever turned ON, the *p*-type substrate is shorted to the source region as indicated in Figure 1.18. As a result of this short, a parasitic diode called the **body diode** exists between the drain and source of the MOSFET. This integral **body diode** can be used in many circuit configurations to advantage, thereby avoiding the use of an external diode that is required in such topologies.

Figure 1.18(a) gives the structural schematic of the MOSFET from a functional viewpoint. However, a typical cross-section of the practical MOSFET is as shown in Figure 1.20. The MOSFET has associated capacitances with respect to its terminals. This parasitic capacitance model of the MOSFET is shown in Figure 1.21. One should be aware that while manufacturing the MOSFET, a reverse diode from source to drain (i.e., anode at source and cathode at drain) gets inherently built by shorting the substrate and the source. This internal diode or the body diode is advantageous as it can be used for freewheeling purposes in inverters that are driving inductance loads.

Referring to Figure 1.20, the capacitance between the drain and the source C_{ds} varies in accordance with the width of the depletion layer that in turn depends on the voltage being supported by the device. The gate source capacitance consists of three components C_{gsn}, C_{gsp} and C_{gsm} (the metal to gate lead capacitance that is not indicated in Figure 1.20). Of these capacitances, C_{gsp} depends on the width of the inversion layer that in turn depends on the applied gate–source voltage. Of particular interest is the feedback capacitance C_{gd}. This capacitance plays a dominant role during switching. This is also the most voltage dependent. C_{gd} is essentially two capacitors in series such that

$$\frac{1}{C_{gd}} = \frac{1}{C_{gdox}} + \frac{1}{C_{gdbulk}} \tag{1.32}$$

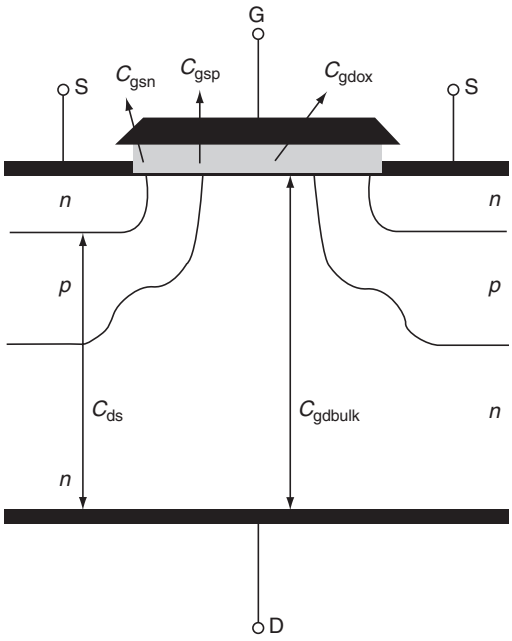


Figure 1.20 | Structure of the MOSFET.

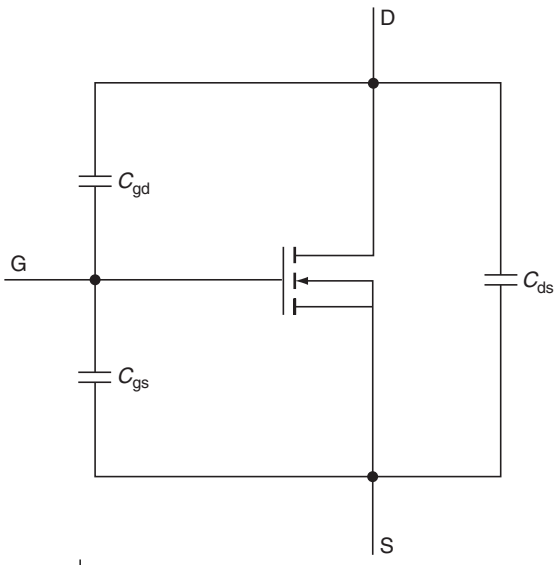


Figure 1.21 | Parasitic capacitance model of the MOSFET.

All the capacitances vary according to the thickness of the depletion region. Figure 1.21 depicts the capacitance model of the MOSFET. Most MOSFET datasheets do not refer to these capacitances. However, they provide information on the input capacitance C_{iss}, the output capacitance C_{oss} and the feedback capacitance C_{rss}. The datasheet capacitances relate to the parasitic capacitances shown in Figure 1.21 as follows:

C_{iss}: Parallel combination of C_{gs} and C_{gd}
C_{oss}: Parallel combination of C_{ds} and C_{gd}
C_{rss}: Equivalent to C_{gd}

Static Characteristics

The magnitude of the gate–source voltage V_{gs} essentially determines the drain current. Figure 1.22(a) shows the drain current i_d versus gate–source voltage V_{gs} characteristics. It can be observed from Figure 1.22(a) that there is a threshold voltage V_{gsT} below which the device is OFF. This V_{gsT} is of the order of 3–4 V in most power MOSFETs. Figure 1.22(b) shows the drain current i_d versus drain–source voltage V_{ds}. In the active region, the drain current is independent of the drain–source voltage and depends only on V_{gs}. This relation is approximately given by

$$i_d \propto (V_{gs} - V_{gsT})^2 \tag{1.33}$$

The MOSFET is said to be in the Ohmic region if V_{ds} is less than $(V_{gs} - V_{gsT})$. The boundary between the Ohmic and active region is depicted in Figure 1.22(b). The Ohmic region corresponds to the saturation region of BJT output characteristics. The MOSFET in the ON-state will operate in the Ohmic region. The drain to source resistance of the MOSFET in the Ohmic region is denoted by $R_{DS(ON)}$ and is an important selection parameter for the device.

In order to reduce the power dissipation in the ON-state, the device should have a low $R_{DS(ON)}$. A V_{gs} of ~10 V will take the MOSFET to the ON-state. However, in practical circuits, a V_{gs} of 15 V is applied in order to take the MOSFET to full enhancement such that the $R_{DS(ON)}$ is low. The maximum V_{gs} that can be applied is limited to +20 V, as only this much V_{gs} can be supported by the thin oxide layer.

Dynamic Characteristics

Power MOSFETs are intrinsically faster than BJTs as they have no excess minority carriers that must be moved into or out of the device during turn-ON and turn-OFF conditions. The only charges that must be

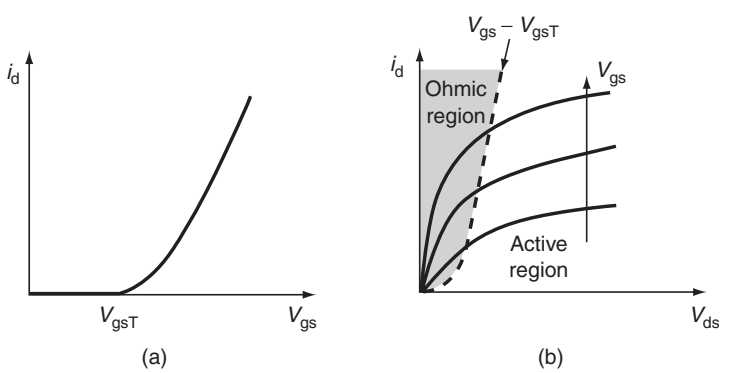


Figure 1.22 | Static characteristics of MOSFET.

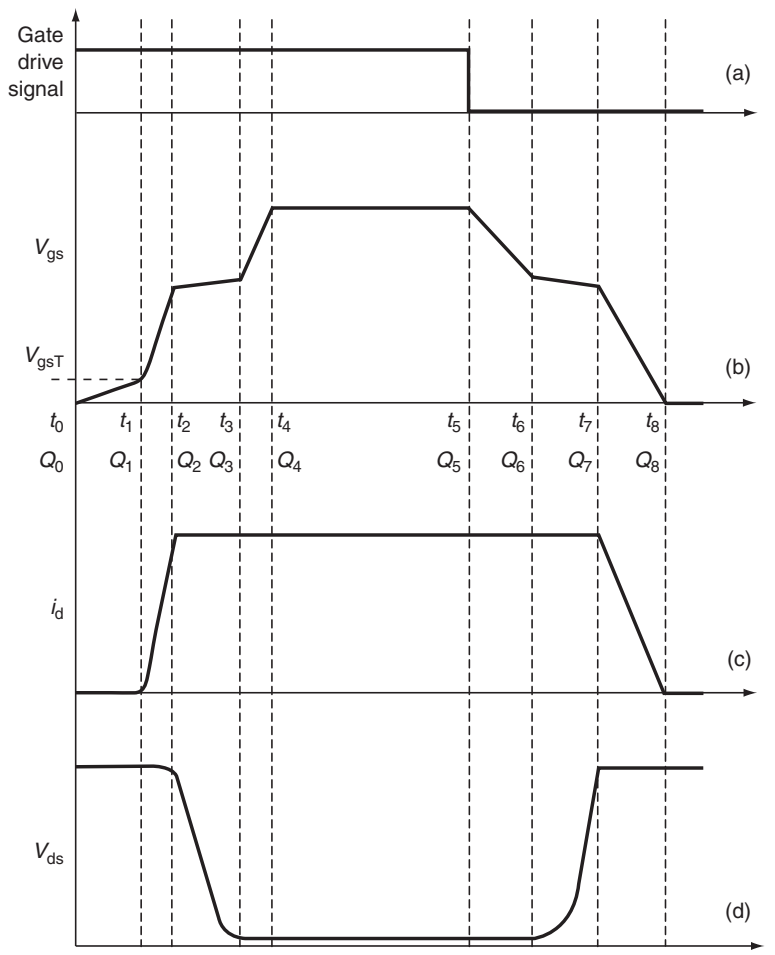


Figure 1.23 | Switching characteristic of MOSFET.

moved in or out are those on the stray parasitic capacitances due to oxide and depletion layers. The parasitic capacitances have been shown in Figure 1.21. The dynamic or the switching characteristics of the MOSFET are shown in Figure 1.23. The switching characteristic of the MOSFET is divided into various intervals. These intervals are explained based on the capacitance model of the MOSFET shown in Figure 1.21. For the purpose of the discussion regarding the switching behavior of the MOSFET, the current fed to the gate can be assumed to be a constant flowing into the gate terminal during the turn-ON period, that is, t_0–t_4. During the period t_4–t_5 when the MOSFET is ON, the gate current is negligible. During the period t_5–t_8, when the MOSFET is turning OFF, the gate current is again assumed as constant but flowing out of the gate terminal. The equivalent circuit of the MOSFET during the turn-ON process at the various intervals is depicted in Figure 1.24. Referring to Figures 1.23 and 1.24, the switching action of the MOSFET is discussed referring to the various time intervals.

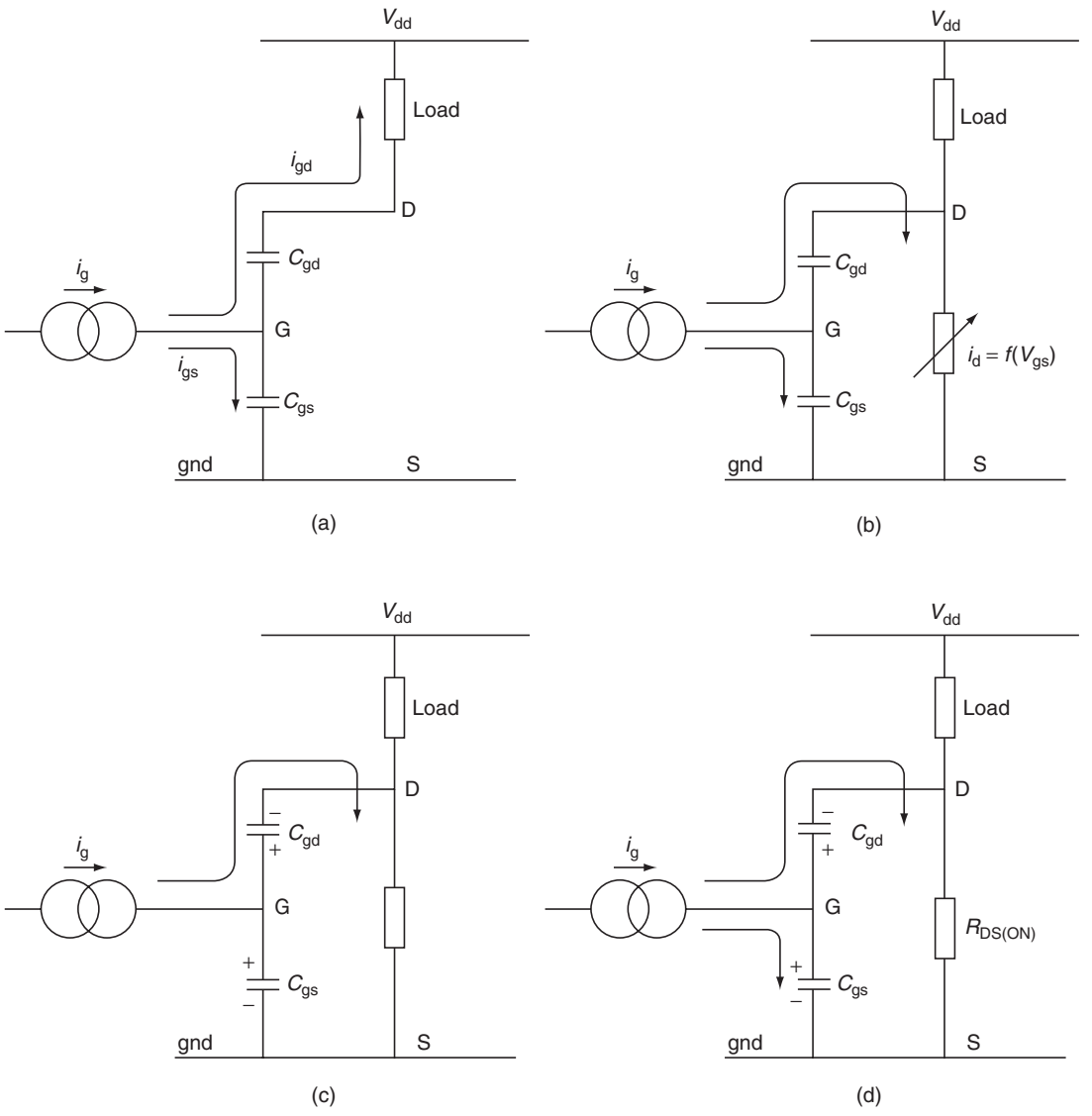


Figure 1.24 Equivalent circuits during turning ON of a MOSFET: (a) During interval $t_0 - t_1$; (b) during interval $t_1 - t_2$; (c) during interval $t_2 - t_3$; (d) during interval $t_3 - t_4$.

During Interval t_0–t_1

At t_0, let there be an initial charge of Q_0 at the gate to source capacitance. An amount of charge $(Q_1 - Q_0)$ is required to charge the gate to the threshold voltage V_{gsT}. The gate current flows through C_{gs} and C_{gd} as shown in Figure 1.24(a).

At turn-ON, C_{gs} is charged by a current

$$i_{gs} = C_{gs} \frac{dV_{gs}}{dt}$$

and C_{gd} is charged by a current

$$i_{gd} = C_{gd}\frac{dV_{gd}}{dt} \approx -C_{gd}\frac{dV_{ds}}{dt}.$$

During Interval t_1–t_2

An amount of charge $(Q_2 - Q_1)$ is required such that the full drain current i_d flows. The equivalent circuit is as shown in Figure 1.24(b).

During Interval t_2–t_3

The equivalent circuit is as shown in Figure 1.24(c). During this interval, almost the entire gate current flows through the capacitance C_{gd}. As a consequence, the charge on C_{gs} is almost constant. This leads to an almost flat V_{gs} waveform during this interval as indicated in Figure 1.23(b). The drain-to-source voltage waveform starts falling as the MOSFET moves further into the Ohmic region. The rate at which V_{ds} falls as C_{gd} gets charged is

$$\frac{dV_{ds}}{dt} = -\frac{dV_{gd}}{dt} = \frac{i_g}{C_{gd}} \tag{1.34}$$

During Interval t_3–t_4

Once the drain–source voltage has dropped to its ON-state value of $i_d R_{DS(ON)}$, C_{gs} gets charged such that the MOSFET is driven to full enhancement. An amount of charge $(Q_4 - Q_3)$ is required for the MOSFET to be taken to full enhancement. The equivalent circuit is as shown in Figure 1.24(d).

During Interval t_4–t_5

During this interval, the MOSFET is fully ON and no charge is required to maintain it in its ON-state. Thus the gate current is zero during the time when the MOSFET is fully ON.

During Interval t_5–t_6

At instant t_5, the gate signal to the MOSFET is made zero. Now the MOSFET begins the turn-OFF process. The charge $(Q_6–Q_5)$ is the excess gate charge from C_{gs} that is required to be removed to bring the MOSFET out of full enhancement.

During Interval t_6–t_7

$(Q_7–Q_6)$ amount of charge has to be discharged to bring the MOSFET out of the Ohmic region to the OFF-state. The drain–source voltage V_{ds} increases during this time as indicated in Figure 1.23. The feedback capacitance C_{gd} is discharged during this period. The rate at which the voltage across the feedback capacitance C_{gd}, reduces will be same as the rate at which the drain–source voltage rises.

During Interval t_7–t_8

$(Q_8–Q_7)$ is the charge required to be discharged from C_{gs} to fully turn-OFF the MOSFET as indicated in Figure 1.23. One should note that the gate drive source should be able to sink during the discharge duration. The interval t_1–t_2 where the drain current rises up to its full value corresponds to the rise time t_r and the interval t_7–t_8 where the drain current falls from its full value to zero corresponds to the fall time t_f. [A typical MOSFET has the following specifications: for example, for BVZ 45 B, 500 V, 10 A MOSFET has C_{gs} ~ 4000 pF; C_{gd} ~ 100 pF; C_{ds} ~ 300 pF; t_r ~ 100 ns; t_f ~ 120 ns.]

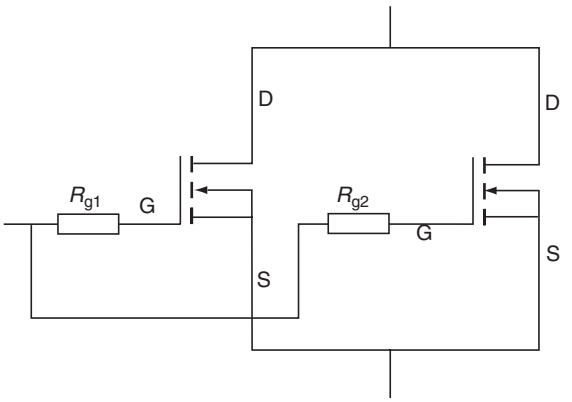


Figure 1.25 | Paralleling of MOSFETs.

Paralleling of MOSFETS

The Ohmic region resistance $R_{DS(ON)}$ is due to majority carriers. It has a positive temperature coefficient. This is advantageous when paralleling the devices. Current sharing resistors in series with the source are normally not required (as in the case of BJTs). Incidentally it can be noted that due to positive temperature coefficient of $R_{DS(ON)}$, the second breakdown will not take place in MOSFETS. In case there is a non-uniform distribution of the drain current, as the current increases in any section, the $R_{DS(ON)}$ will increase for that particular local section and as a consequence the current will reduce in that section and spread from the hot spot region to other areas.

If MOSFETS have to be connected in parallel, the schematic shown in Figure 1.25 is generally used. The drains and sources can be connected in parallel directly. While this can take care of the static condition, it is necessary to put small gate resistances of 10–100 Ω to damp high frequency oscillations that might result between gate leads and the parasitic device capacitances.

One should note that for a given rated current, the voltage drop will be much higher for a high-voltage MOSFET ($R_{DS(ON)}$ ~1 to 2 Ω) than for a low-voltage device (R_{DS} ~ 0.1 Ω). This is due to the fact that the bulk *n*-region, that is, the drain region has to be larger in order to be capable of withstanding the electric field due to the higher applied voltage.

Handling Precautions for MOSFET

The gate–source oxide layer is very thin. Therefore, when touched with our body (that can have a minimum of 1–2 μC charge) the fields created can be extremely high so as to breakdown the oxide layer.

Remarks on BJT and MOSFET

1. BJTs have lower conduction losses in ON-state but higher switching losses whereas power MOSFETs have higher conduction losses and lower switching losses. Thus at higher frequencies, power MOSFETS are the obvious choice.
2. BJTs are current-controlled devices whereas MOSFETs are voltage-controlled devices.
3. It is easier to parallel power MOSFETs when compared to BJTs.
4. BJTs suffer from secondary breakdown, but the power MOSFETs do not suffer from secondary breakdown.

EXAMPLE 1.1

Consider a MOSFET with a C_{gs} = 4000 pf. From the basic charge–potential relationship,

$$Q = CV$$

The voltage across the gate–source capacitance that would get developed for a charge of 1 μC is

$$V_{gs} = \frac{Q}{C_{gs}} = \frac{1\times10^{-6}}{4\times10^{-9}} = 250 \text{ V}$$

The maximum gate–source voltage that is allowable is only 20 V. Therefore, just touching a MOSFET may damage it permanently and hence the MOSFETs must be always carried in a conducting pad, so that the charge leaks if it tries to build up. While not in use, it is recommended to short the gate and source of the MOSFET with a conducting jumper or wire so that the gate–source voltage does not build due to spurious electrostatic fields. In order to limit the V_{gs} to 20 V, it is always a good practice to clamp the voltage across the gate-source terminals of a MOSFET by connecting a 15 V Zener diode close to gate and source leads.

Power Dissipated in a Power MOSFET

The ON-state loss for the MOSFET is given by

$$P_{ON} = I_{drms}^2 R_{DS(ON)} \tag{1.35}$$

$R_{DS(ON)}$ has a temperature coefficient of approximately 0.6% per °C and should be considered to calculate $R_{DS(ON)}$ (and P_{ON}) as the device temperature rises. In a manner similar to the discussion with regard to the BJT, the switching loss for resistive loads, assuming linear rise and decay of V_{ds} and i_d, is given as

$$P_{switching} = \frac{1}{6} V_{ds} I_d (t_r + t_f) f_s \tag{1.36}$$

Model for Power MOSFET

There are different levels of MOSFET models that have been used depending on the application. For integrated circuit fabrication using simulation, more complex models are necessary. Figure 1.26 shows a simple model that can be used for macro circuit simulation. In Figure 1.26, the capacitance subscript "gs" means gate to source, "gd" means gate to drain, "gb" means gate to bulk and "sb" means substrate to bulk.

1.10 Insulated Gate Bipolar Transistor (IGBT)

The IGBT is a combination of the advantageous features of MOSFET (like high drive input impedance, voltage control, fast switching) and BJTs (like low ON-state losses and high OFF-state voltage capability). The IGBT is also a three-terminal device. It has an emitter lead E (analogous to the emitter of

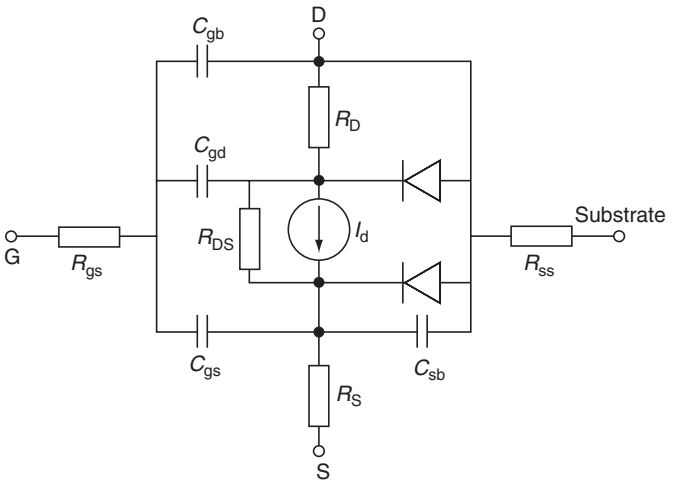


Figure 1.26 | Simulation model of a MOSFET.

the BJT), a collector lead C (analogous to the collector of the BJT) and a gate lead G (analogous to the gate of the MOSFET). The flow of current from the collector to the emitter is controlled by the voltage applied between the gate–emitter terminals. Therefore, like the MOSFET, the IGBT is also a voltage-controlled device. In the ON-state, the IGBT can allow current in only one direction (from collector to the emitter for *n*-channel IGBTs and from emitter to the collector for *p*-channel IGBTs). Further, they can support bipolar voltages during the OFF-state.

A structural schematic of an *n*-channel IGBT is depicted in Figure 1.27(a). It consists of four layers: a *p*-type semiconductor that is connected to the collector C, an *n*-type semiconductor that is connected to the emitter E and in between is a *p*–*n* semiconductor junction. The gate is connected to a metallic conductor. This is insulated from the bulk of the IGBT by an insulator that is generally an oxide of metal, like, silicon dioxide. The symbol for the *n*-channel IGBT is shown in Figure 1.27(b) and that for a *p*-channel IGBT is shown in Figure 1.27(c).

When a positive voltage is applied at the gate as indicated in Figure 1.28, an electric field will be directed through the insulating oxide layer. This brings about polarization of charges within the oxide layer. As a consequence, electrons from the *p*-material get attracted towards the insulating oxide layer thereby creating an induced *n*-channel that bridges the two *n*-type sections of the IGBT. The induced *n*-channel along with the two *n*-sections forms an equivalent *n*-type semiconductor. The *p*-type semiconductor connected to the collector C along with the equivalent *n*-type semiconductor forms an equivalent *p*–*n* junction and behaves like a diode. Because of the formation of the equivalent *p*–*n* junction, the IGBT is capable of withstanding **reverse** voltages also during the OFF-state.

From the structural schematic of the IGBT, it is evident that an IGBT can be viewed equivalently as a MOSFET with a diode connected to its drain as indicated in Figure 1.29. Alternatively, it can also be viewed as a MOSFET and a PNP transistor as shown in Figure 1.30.

The detailed structure of the IGBT is shown in Figure 1.31. One should note that the IGBT has a parasitic thyristor between its collector and emitter as indicated in Figures 1.31 and 1.32. This parasitic thyristor should not latch. If it latches then the IGBT will loose the gate control. Therefore, when the IGBT is

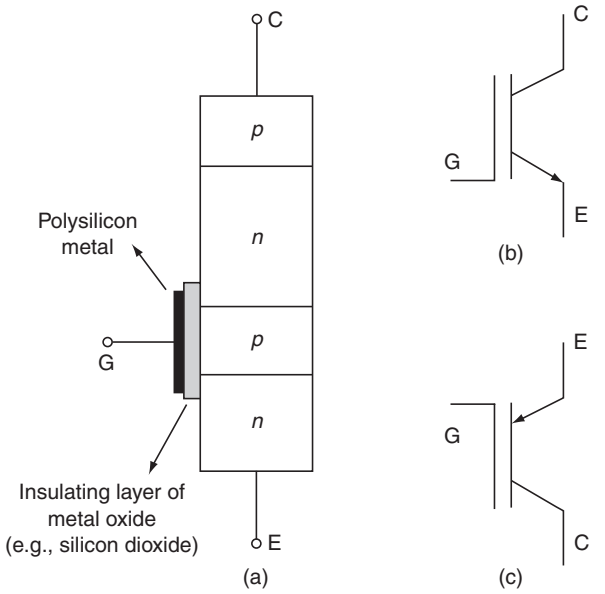


Figure 1.27 (a) Structural schematic of an *n*-channel IGBT; (b) symbol of an *n*-channel IGBT; (c) symbol of a *p*-channel IGBT.

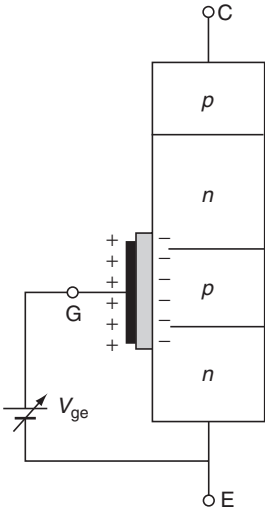


Figure 1.28 Operation of the IGBT.

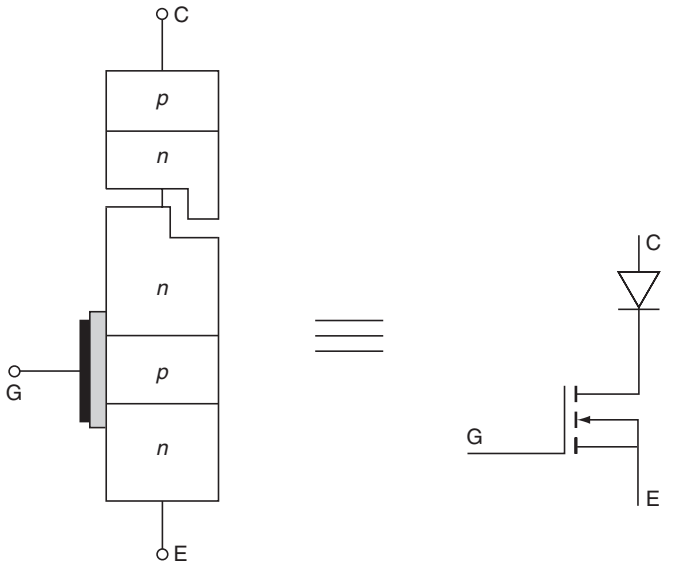


Figure 1.29 | Equivalent circuit of IGBT: Diode and MOSFET.

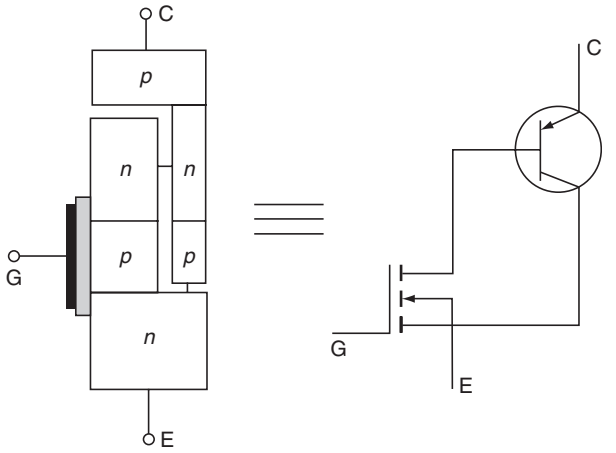


Figure 1.30 | Equivalent circuit of IGBT: Transistor and MOSFET.

manufactured, R_s as indicated in Figure 1.32 is made very low so that the drop across it is less than cut-in voltage of the base–emitter of the NPN transistor in the parasitic thyristor. This way the NPN transistor is always off, thereby avoiding the latching problem. One should be aware that while manufacturing the IGBT, a reverse diode from emitter to collector (i.e., anode at emitter and cathode at the collector) is built separately as it is not inherent like in the case of the MOSFETs. This internal diode or the body diode is required for freewheeling purposes in topologies that are driving inductance loads.

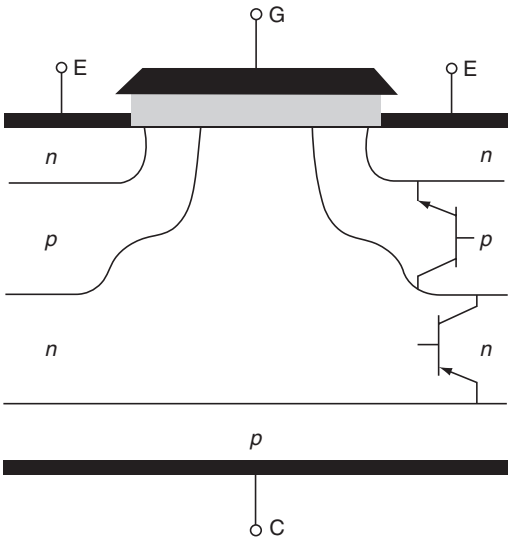


Figure 1.31 | Structure of an IGBT.

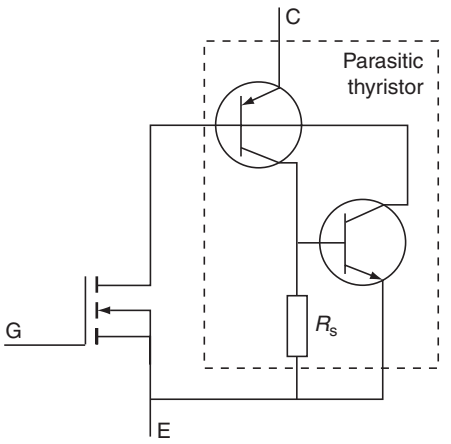


Figure 1.32 | Equivalent circuit of the IGBT showing the parasitic thyristor.

Static Characteristics

The i_c–V_{ce} characteristics of an *n*-channel IGBT are shown in Figure 1.33 and the i_c–V_{ge} characteristics are shown in Figure 1.34. From Figure 1.33 it is observed that the IGBT can support both forward and reverse voltages during the OFF-state. From Figure 1.34, it is seen that if V_{ge} is less than the threshold voltage, the IGBT is in the OFF-state.

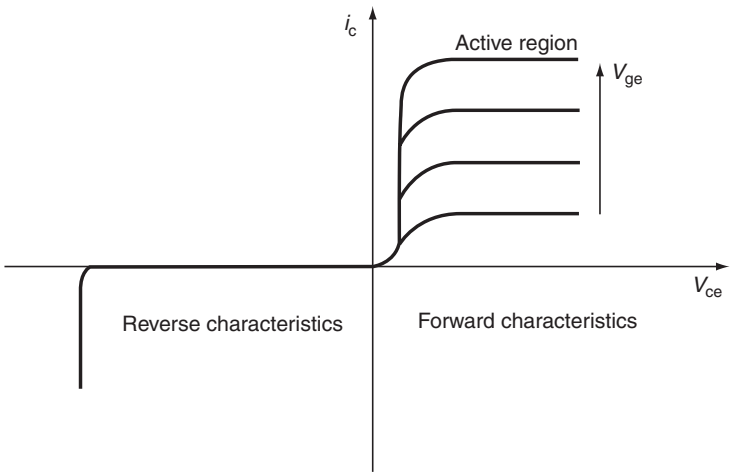


Figure 1.33 | Static i_c–V_{ce} characteristic of an IGBT.

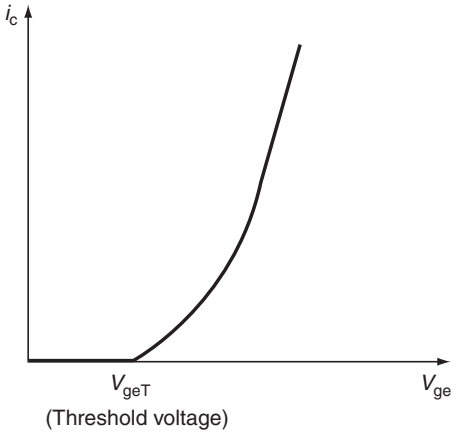


Figure 1.34 | Static i_c–V_{ge} characteristic of an IGBT.

In comparison, for a given i_c, the $V_{ce(sat)}$ of the IGBT is less than the V_{ds} of a comparable MOSFET when ON but greater than $V_{ce(sat)}$ of a comparable BJT.

Dynamic Characteristics

The dynamic behavior of an IGBT is similar to that of a MOSFET when viewed from the gate side and is similar to that of a BJT when viewed from the collector side. However, the maximum switching frequency is limited as compared to that of a MOSFET.

The switching times of the IGBT are related to the gate–emitter voltage and collector-current waveforms. Figure 1.35 shows the typical switching waveforms. The switching times of an IGBT are mainly determined by its internal capacitances and parasitic inductances together with the internal resistance

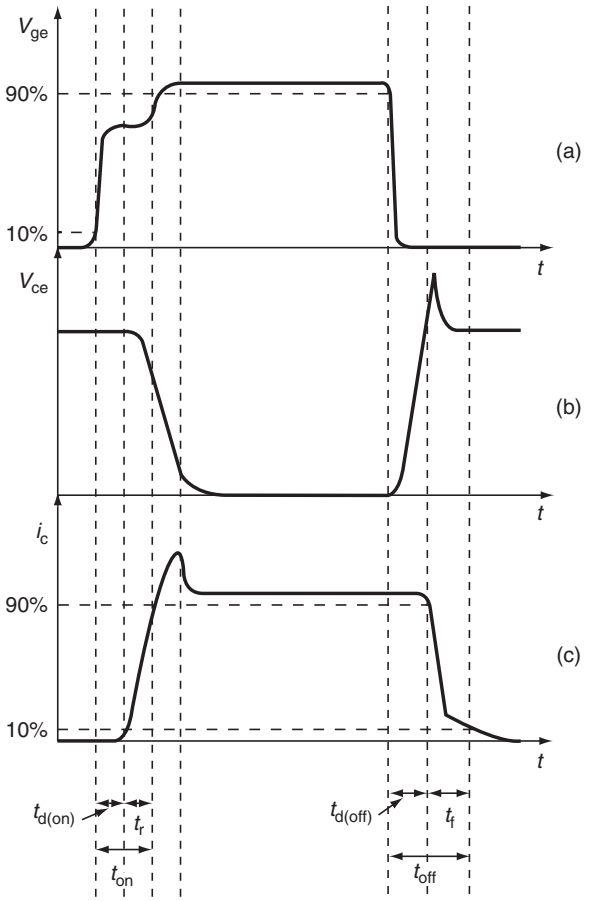


Figure 1.35 | Dynamic switching characteristics of an IGBT.

of the gate-side voltage source. In order to charge and discharge the capacitances rapidly and to reduce the transients caused by the gate circuit inductance, a low internal impedance of the gate-side voltage source is desirable. It shortens the switching times and reduces the switching losses. On the other hand, a very fast turn-ON causes a high peak reverse recovery current through the body diode that appears as an additional peak collector current. During very fast turn-OFF, a high transient voltage is caused by the parasitic collector–emitter inductance as shown in Figure 1.35(b).

It is very important to keep the parasitic inductance in the gate circuit at a minimum by using very short leads. This inductance might otherwise generate parasitic oscillations in conjunction with the IGBT capacitances. The maximum rated gate–emitter voltage as specified in the datasheets in most cases is ±20 V. It is recommended to connect a 15 V Zener diode between the gate and the emitter as a protective clamper.

Turn-ON

During turn-ON, the IGBT behavior is more like that of a MOSFET. The collector–emitter voltage, the collector-current and the gate–emitter voltage waveforms are similar to the turn-ON waveforms of the

MOSFET. When the gate–emitter voltage reaches the threshold value of V_{geT}, the collector-current starts to rise. The time interval between the instant when V_{ge} reaches 10% of its final value and the instant when i_c reaches 10% of its final value is called the turn-ON delay time $t_{d(on)}$. The subsequent time interval up to the instant when the collector current reaches 90% of its final value is called the rise time t_r. During this period of time, most of the turn-ON power dissipation takes place. The sum of the turn-ON delay time $t_{d(on)}$ and the rise time t_r is called the turn-ON time t_{on}. At the end of t_{on}, the collector–emitter voltage V_{ce} often has not yet fallen to its final value of $V_{ce(sat)}$. This has to be considered when calculating the turn-ON dissipation. The collector-current peak shown in Figure 1.35 indicates the peak reverse recovery current of the body diode or the freewheeling diode. This peak current has to be taken into account in any turn-ON power dissipation calculations as well.

Turn-OFF

During turn-OFF, the IGBT behaves more like a BJT. To turn the IGBT OFF, the voltage in the gate control circuit is switched to zero. Referring to Figure 1.35, there is first a turn-OFF delay time $t_{d(off)}$ that is the interval between the instant when the gate–emitter voltage has fallen to 90% of its initial value and the instant when the collector current has fallen to 90% of its initial value. The subsequent period of time up to the instant when the collector current has fallen to 10% of its initial value is called the fall time t_f. The sum of $t_{d(off)}$ and t_f is called the turn-OFF time t_{off}.

Tail Current

The tail current and tail time are properties that are specific to the IGBT. However, they also depend on the operating conditions. The tail current is higher for lower saturation voltage $V_{ce(sat)}$. The trade-off for the favorable reduction of the conducting state power loss is an increase in the tail current. Since at pulse frequencies of 10–20 kHz the switching losses are more prevalent, most IGBTs are designed for a low tail current at the expense of the saturation voltage which would be a little higher than the minimum possible value.

Remarks on IGBT

1. The power loss calculations for the IGBT are similar to the loss calculations indicated for BJTs and MOSFETs.
2. The gate drive power is low such as in the case of MOSFETs.
3. IGBT has considerably greater overload capability as compared to a MOSFET. For a MOSFET, the peak overload is about 5–6 times the continuous drain current whereas for the IGBT, the peak overload is about 20 times the continuous collector current. An IGBT is capable of taking a short-circuit current (across a 600 V bus) for a period of 10 μs.
4. Over voltage robustness is less than that of MOSFETs.
5. IGBTs are more like the BJTs during turn-OFF and like the MOSFETs during turn-ON.
6. Unlike the BJT, where the collector current is limited by the current gain, that is ($h_{FE(sat)}i_b$), the IGBTs do not have such limitations. The collector current is limited only by the external source.

1.11 Thyristor Family

The term thyristor is a generic name for a semiconductor switch having four or more layers and is in essence a *p–n-p–n* structure. Thyristors form a large family of semiconductor switches. If an Ohmic connection is made to the first *p*-region and the last *n*-region and no other connection is made to any other intermediate region, then the device is a diode thyristor. If an additional Ohmic connection is made to the

intermediate *n*-region or the intermediate *p*-region, the device is called a triode thyristor. If an Ohmic connection is made to both intermediate regions, then the device is a tetrode thyristor. All such devices have a forward characteristic of the form shown in Figure 1.36.

There are three categories of thyristor reverse characteristic: blocking (as in normal diodes like in symmetric SCRs), conducting (large reverse current at low reverse voltages like in asymmetric SCRs and GTOs) and approximate mirror image of the forward characteristic (bi-directional thyristors like in DIACs and TRIACs). The simplest thyristor structure and the most common is the reverse blocking triode thyristor usually referred to as the silicon-controlled rectifier (SCR). The more complex thyristor structure is the bi-directional triode thyristor or TRIAC. In this section, the various devices in the thyristor family will be briefly reviewed.

Silicon–Controlled Rectifier (SCR)

SCR is a three-terminal device with an anode terminal A, a cathode terminal K and a control or gate terminal G. Structurally it is similar to a BJT with an extra *p*-layer at the anode. The structure of the SCR is depicted in Figure 1.37(a) and the circuit symbol is shown in Figure 1.37(b).

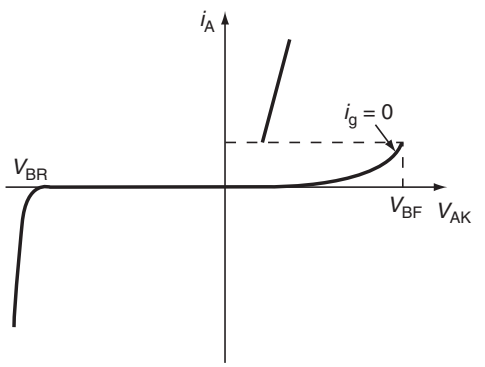


Figure 1.36 | Forward characteristics of the thyristor family.

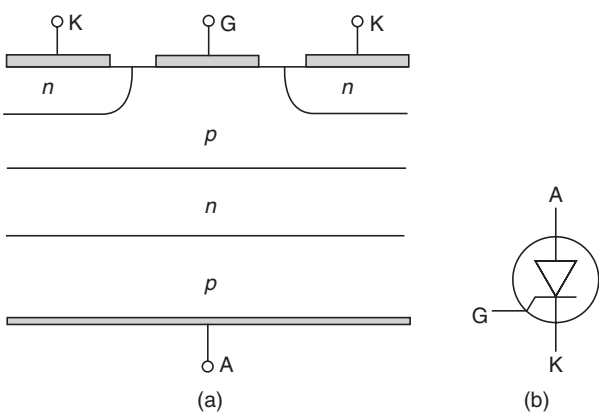


Figure 1.37 | (a) Structure of an SCR; (b) symbol of an SCR.

Static Characteristics

Application of a negative voltage to the anode–cathode of an SCR will reverse bias the SCR. In this condition, the junctions J_1 and J_3 are reverse-biased and J_2 is forward-biased. The device is said to be in the OFF-state and blocks the reverse voltage. On the other hand, applying a positive voltage to the anode–cathode, as shown in Figure 1.38, with the gate open, will forward bias the SCR. With the anode–cathode voltage positive, it can be observed that junctions J_1 and J_3 are forward-biased and J_2 is reverse-biased. The SCR is still in the OFF-state and blocks the applied forward voltage. This is a property unique to the thyristor family. If forward-bias voltage is further increased, the junction J_2 will breakdown and the device turns ON behaving in a manner similar to a diode. The static characteristic is shown in Figure 1.39. Figure 1.39(a) shows the static characteristics when the gate current is zero. The voltage at which the junction J_2 breaks down and the SCR conducts is called the forward breakover voltage V_{BF}. If the gate of the SCR is connected to a voltage source and a gate current is allowed to flow, then the breakover voltage reduces as indicated in Figure 1.39(b).

The SCR action is best understood by modeling it as two BJTs connected as shown in Figure 1.40. With the gate open or shorted to the cathode, the device is OFF and no current flows from anode to cathode, except for a very low leakage current. If an external positive current pulse is applied to the gate, it becomes the base current of the NPN transistor Q_2. Consequently, the collector current of Q_2 supplies the base current to the PNP transistor Q_1. The collector current of Q_1 then further increases the base current of Q_2 and so on. This iterative action maintains the device in the conducting state even if the gate signal is now removed. The device continues to conduct till the anode voltage is less positive than the cathode voltage. Figure 1.39(b) shows the static characteristics with gate current as a parameter. It should be noted that once the device is ON, the gate looses complete control of the device and the device cannot be turned OFF through the gate. The SCR can be switched OFF only if the anode current is brought below a threshold called the **holding current** value for that SCR.

Dynamic Characteristics

di/dt Effect On supplying the gate trigger to a forward-biased SCR, the SCR will turn ON. The anode current starts to rise after a small delay. The rate of change of the anode current will depend on the nature of

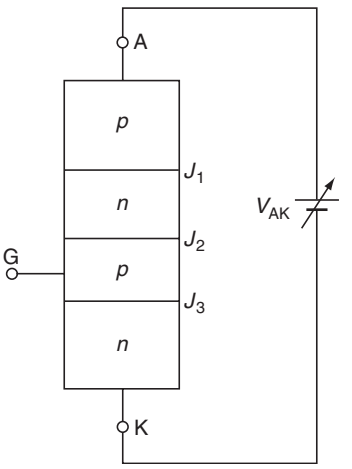


Figure 1.38 | Biasing the SCR.

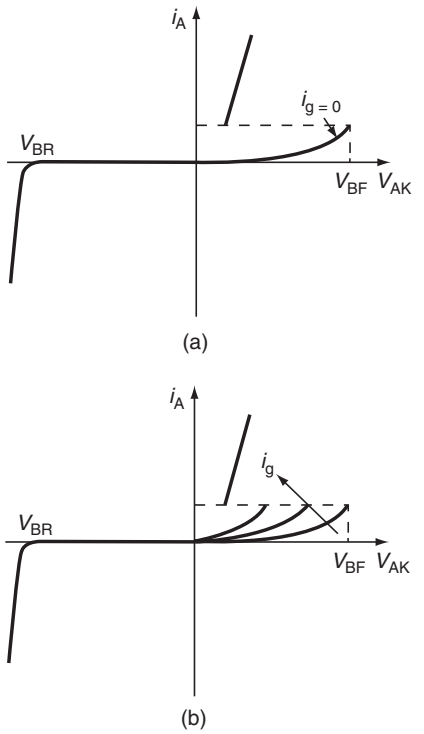


Figure 1.39 | *V–I* characteristic of an SCR (a) with gate open; (b) with gate current applied.

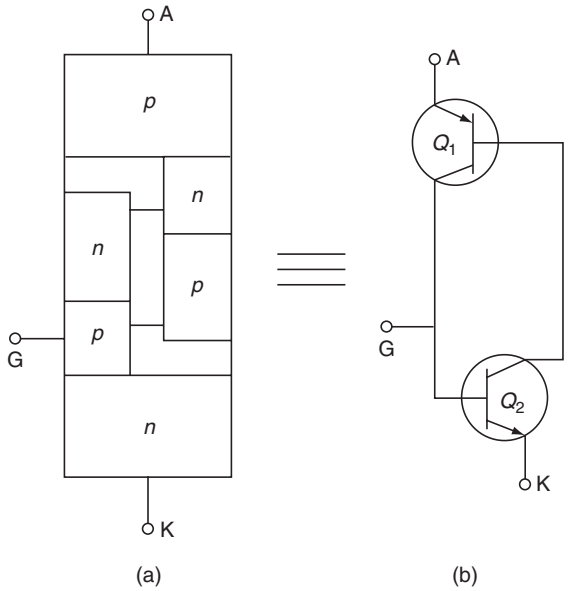


Figure 1.40 | Two transistor equivalent circuit of an SCR.

the load. Initially the ON-state current is concentrated in a small area around the gate region. If the rate of rise of this ON-state current is high, then the area around the gate region will become overheated that may permanently damage the device. Therefore, it should be ensured that the external circuit di/dt encountered by the SCR should be less than the rated di/dt of the device.

Turn-OFF Characteristics The turn-OFF characteristic waveforms of the anode current and the anode–cathode voltage are similar to that of the diode with the added restriction that all junctions including the gate junction should fully recover before the SCR gets into a forward-blocking state. The turn-OFF characteristics are depicted in Figure 1.41.

A high rate of rise of the applied forward voltage across the anode and cathode (dv/dt) can trigger the SCR. To keep the SCR turned OFF, the re-application of the OFF-state forward voltage must be delayed to avoid re-triggering due to high dv/dt. The circuit commutated recovery time t_q is measured from the zero crossover of the current as shown in Figure 1.41. It should be noted that t_q is an important parameter. SCRs are often classified as line-commutated or converter grade if $t_q > 50$ μs and as inverter grade if $t_q < 20$ μs. For the line commutated converter, t_q is less important since the half-period is 10 ms. But for inverter circuits switching at higher frequencies, inverter grade SCRs with appropriately lower t_q will have to be used.

dv/dt Effect Figure 1.41 shows a possible problem with large rate of forward voltage, dv/dt. If dv/dt is large, the SCR may self-trigger even in the absence of gate current. This is due to the fact that there exists a leakage current through the reverse-biased junction. This leakage current flows through the capacitance of junction J_2 and is given by $i = c$ dv/dt. This leakage current is dependent on dv/dt. If this leakage current

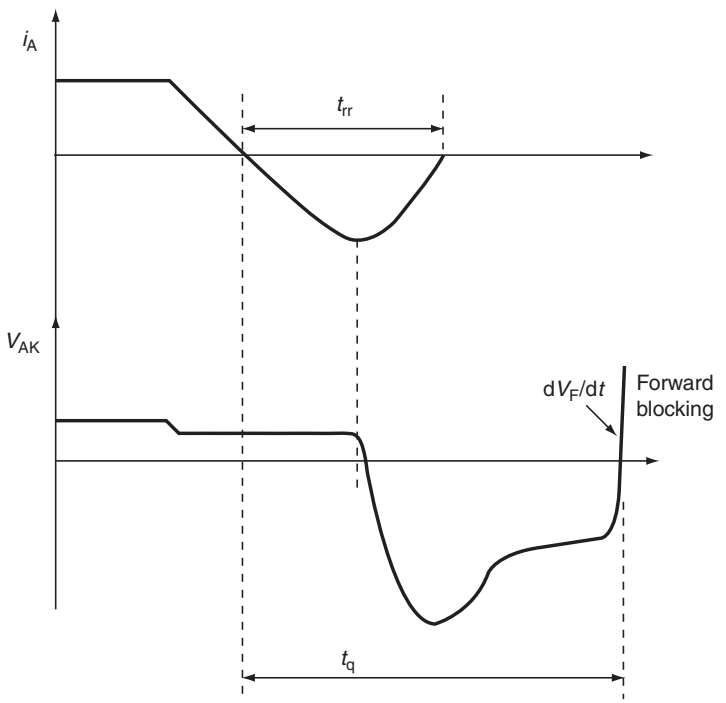


Figure 1.41 | Anode current and anode–cathode voltage during turn-OFF.

is larger than the latching current of the device, then the device will self-trigger and go into conduction even in the absence of the gate current. Therefore a high dv/dt has the possibility of undesirable latch up of the device. Practical SCRs have dv/dt limits of about 200–500 V/μs. In circuits wherein the dv/dt exceeds the rated dv/dt limits of the device, additional circuits called snubber circuits will have to be used to lower the dv/dt.

Other Parameters

The various current ratings like I_{Tav}, I_{Trms}, I_{Tsurge} are similar to those discussed for the diode. The ratings are a bit complicated compared to a diode. The value given in the datasheet for the mean ON-state current I_{Tav} is valid for a certain waveform and case temperature. I_{Trms} signifies the heating effect due to i^2R dissipation and is limited due to the thermal stress on the device. I_{TSM} is the maximum permissible peak current of half sine wave with a duration of 10 ms at a specified temperature.

Power dissipated in a thyristor is calculated in the same way as that of the diode. Gate-current losses add to the power losses in the case of thyristors as compared with the diode losses. Power loss calculation depends on the type of waveform too. Thyristors are very rugged devices as they can carry considerable overload currents without exceeding the junction temperature.

Circuit Model

There are many circuit models of SCRs available in the literature and circuit simulator libraries like SPICE. One such simple model is given here. It can be constructed using the two-transistor analog of an SCR as shown in Figure 1.42. Diode DFOR is added to model the forward breakdown. A resistance is added across the base–emitter of Q_2 to provide a discharge path for the stored charges in Q_2. The various parameters for the PNP and NPN transistors are included in the model to give a satisfactory simulation performance.

DIACs and TRIACs

DIAC is a two-terminal device without any gate. It is a bi-directional switch. The breakdown voltage is generally kept low (~30 V) because it is generally used as a threshold device in many applications. Figure 1.43(a) shows the DIAC characteristics and Figure 1.43(b) depicts its symbol.

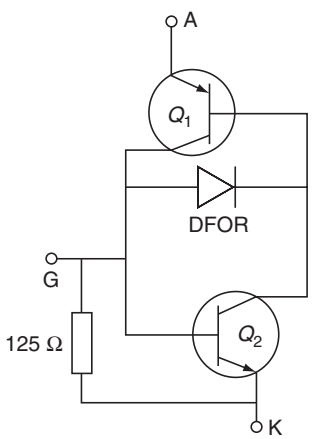


Figure 1.42 | Circuit model of an SCR.

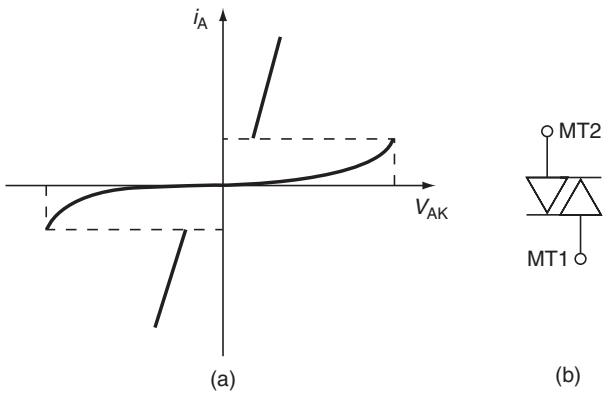


Figure 1.43 | DIAC: (a) *v–i* characteristic; (b) symbol.

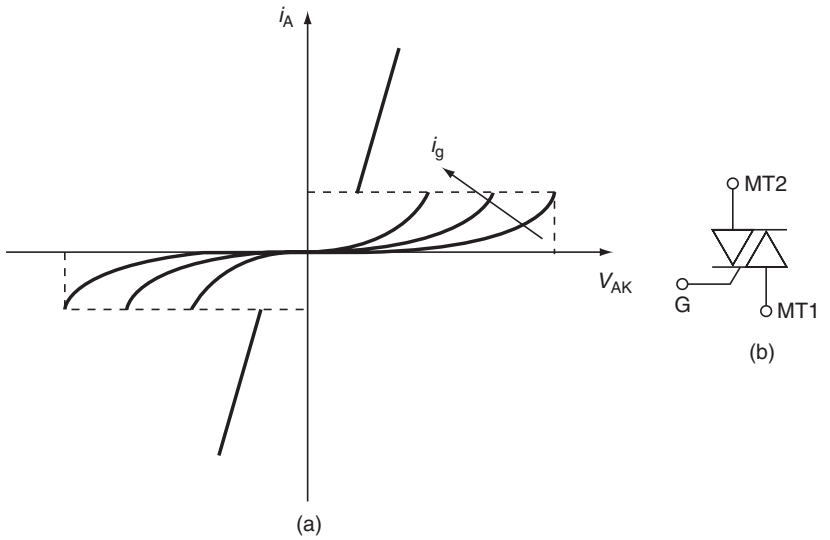


Figure 1.44 | TRIAC: (a) *v–i* characteristic; (b) symbol.

TRIAC is also a bi-directional thyristor switch but it has a gate terminal that is used for controlling the turn-ON of the TRIAC. It is mainly used for AC control applications. As the switch is bi-directional, its terminals are called main terminal 1 and 2 (MT1 and MT2), instead of anode and cathode. Figure 1.44(a) shows the characteristics of the TRIAC and Figure 1.44(b) depicts its symbol. A simple TRIAC control circuit is shown in Figure 1.45. This circuit uses a DIAC as a threshold device for triggering the TRIAC at a particular source phase angle.

There are different ways in which the gate can be fired. The most common is with reference to MT1. The polarity of voltage across MT2 and MT1 decides the direction of current flow, and gate pulse is always positive irrespective of direction of current flow. TRIACs are generally used for 50 Hz applications. With inductive load, the TRIAC will be subjected to large dv/dt and therefore protective snubber circuits must be used.

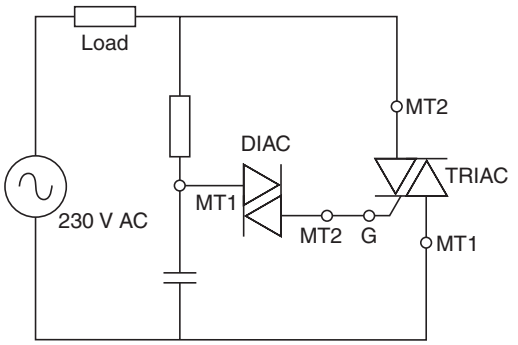


Figure 1.45 | TRIAC control circuit using DIAC.

Gate Turn-OFF Thyristor (GTO)

It can be recalled that an SCR cannot be turned OFF by gate control. It needs an extra turn-OFF/commutation circuit. GTO has been devised to provide turn-OFF through the gate terminal. The GTO is a four-layer device that can be turned ON with a positive gate current pulse and turned OFF with a negative gate current pulse. Figure 1.46(a) shows some of the commonly used GTO symbols.

The physical structure of a GTO is shown in Figure 1.46(b). Referring to figure, it may be observed that the gate and cathode are interspersed. Such a sectionalized geometry helps in having better gate control. The anode layer has *n*- and *p*-regions interspersed as shown in Figure 1.46(b). This gives a structure termed as the "*anode short*" structure that is found to improve the turn-OFF time.

GTO action can also be described by the two-transistor model as discussed for the case of an SCR. By suitably controlling the doping profiles of the *n*- and *p*-regions, the α and β of the transistors, it is possible to achieve gate turn-OFF with a negative gate current. Removal of excess charge is an important part of the turn-OFF process. The negative gate current and the anode current during the turn-OFF process are depicted in Figure 1.47.

The SCR and GTO differences are highlighted below:

1. GTO has higher ON-state drop as compared to the SCR. The GTO has an ON-state drop of about 3–4 V whereas the SCR has an ON-state drop of around 1.5–2 V.
2. The latching and holding currents are high for a GTO. As an example, a typical 110 A SCR has I_H of 150 mA and I_L of 300 mA, whereas even a low current GTO, a typical 18 A GTO has $I_H = 600$ mA and $I_L = 900$ mA. For 30 A GTO, they are more than 1 A. Therefore the triggering pulse must be long enough or continuous for the GTO unlike the case of the SCR.
3. The reverse-blocking capability of a GTO is very low (~20 V) and hence cannot be used in rectifier applications.
4. When the device is to be turned OFF, the turn-OFF amplification factor (anode current with respect to the turn-OFF gate current) can be anywhere between 1 and 5 and the turn-OFF time between 1 and several milliseconds. There is a maximum current rating beyond which gate cannot turn OFF the GTO. This is called repetitive controllable ON-state current (or extinguishable current).
5. GTO has a large dv/dt capability (1000 V/μs) compared to a SCR (200–500 V/μs).
6. GTOs can be used up to 1–4 kHz of switching frequency range.

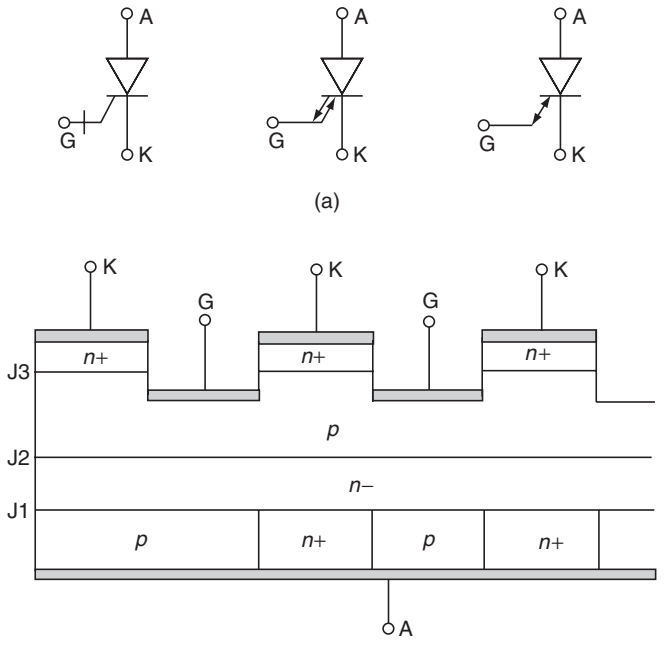


Figure 1.46 | (a) Symbols of GTO; (b) structure of a GTO.

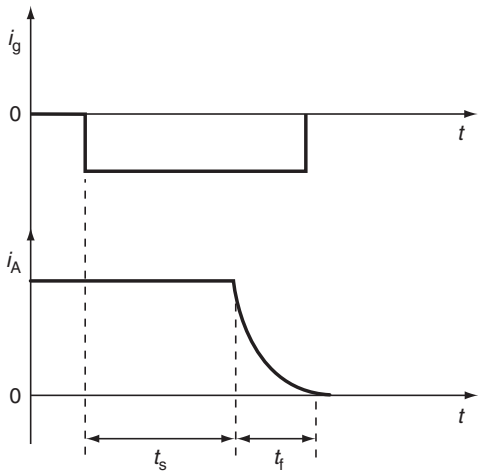


Figure 1.47 | Turn-OFF characteristics of a GTO.

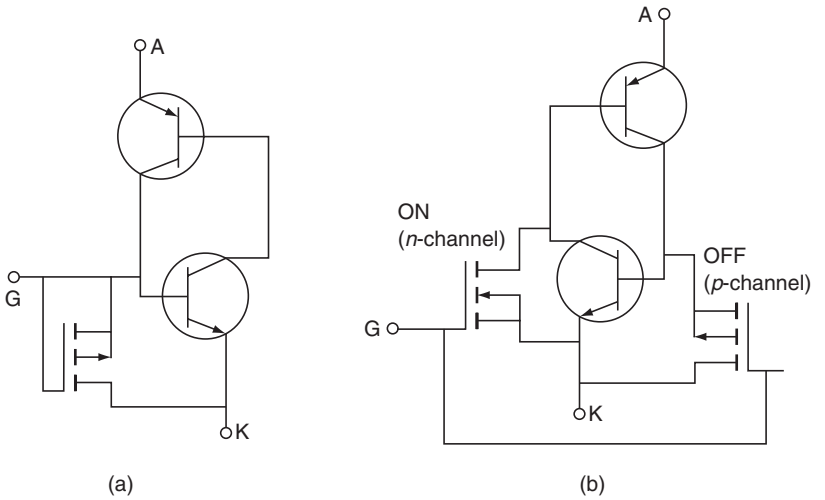


Figure 1.48 | MCT configurations.

MOS-Controlled Thyristor (MCT)

MCT is an interesting device wherein both the turn-ON and turn-OFF can be controlled. However, commercially it is not as popular as the MOSFETs and the IGBTs. Figure 1.48 shows some possible MCT functional configurations. Figure 1.48(a) shows an MCT where turn-ON is done in the usual way but turn-OFF is done using a MOSFET. Figure 1.49(b) shows another configuration wherein an *n*-channel MOSFET is used for turning ON the MCT and a *p*-channel MOSFET is used for turning OFF the MCT. The major features of an MCT are:

1. It has low conduction voltage drop: 1/3 to 1/2 of a comparable IGBT.
2. It combines high-speed, low-power gate turn-ON–turn-OFF advantages.
3. It has high dv/dt and di/dt capabilities.

CONCLUDING REMARKS

This chapter discussed few important viewpoints of the semiconductor devices giving a flavor for the operation of the various devices as a power switch. It should, however, be observed that no single switch meets all the features of an ideal switch as described in Section 1.1. One should be aware of the non-idealities of each type of the power switch so that the power electronic circuits can be modeled with greater closeness to the physical system.

There are many semiconductor devices available commercially that have been fabricated with newer and improved technologies. However, this chapter discusses only a few generic types, highlighting the important issues that need to be studied and understood when encountered with a new device. With special regard to the semiconductor devices being used as power switches in power electronic circuits, one should study with emphasis on

three aspects in general, viz. (a) static characteristics that give insight on the steady-state operating points, (b) dynamic characteristics that give the behavior of the switch during turn-ON and turn-OFF and (c) power loss within the device due to conduction and switching.

The theoretical insights should be cemented with the experience that can be gained from practical work to strengthen the understanding of the device. The next section provides some laboratory exercises that should be implemented both on a simulation platform like spice and by hardware breadboarding to appreciate the functional features of a particular power semiconductor switch. The laboratory exercises are focused on popular devices like diodes, BJTs, MOSFETs and IGBTs. However, the exercises can be extended for any other power semiconductor switch too.

LABORATORY EXERCISES

1. Consider the diode test circuit shown in Figure 1.49 where V_i is a voltage source and D is a diode. The diode D is appropriately chosen such that the current and voltages are within ratings.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Hardware breadboarding

Tasks for study:

(a) Use a variable DC voltage source. Measure the current through and the voltage across the diode for different input DC voltage values. Reverse the polarity of the input voltage source and measure the current through and the voltage across the diode for different input voltage values. Tabulate the diode current and voltage values and plot the *i*–*v* characteristics of the diode.

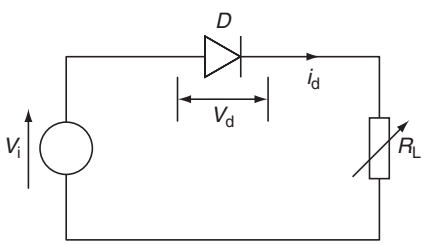


Figure 1.49 | Diode test circuit.

(b) Find out the forward dynamic resistance of the diode from the *i*–*v* characteristic.

(c) Use a square waveform input voltage source and observe the current through the diode and the voltage across it during turn-ON and turn-OFF.

(d) Compute the experimental switching loss and compare with the theoretical estimate.

(e) Compute the experimental conduction loss and compare with the theoretical estimate.

(f) Repeat steps (a)–(c) for different loads by varying R_L.

(g) Repeat steps (a)–(c) for different source frequencies. What is the effect on the diode power loss?

(h) Measure the t_{rr} of the diode from the current waveform through the diode. What is the effect of the load on t_{rr}?

(i) Measure the reverse recovery charge Q_{rr} from the experimental results.

(j) Use a sinusoidal waveform input voltage source and observe the current through and the voltage across it during voltage transition from positive to negative. Compute the power loss in the device and compare the loss with that obtained when using a square waveform input source.

2. Consider the two BJT test circuits shown in the Figure 1.50 where V_b is the base drive pulse source. V_b should be chosen such that it is capable

of both sourcing and sinking current. V_c is the collector DC supply voltage. R_c is the collector load and R_b is the base drive resistor. The BJT is chosen with appropriate rating to handle the collector current and the maximum V_{ce} as decided by R_c and V_c, respectively. The circuit of Figure 1.50(a) consists of a resistive collector load and that of Figure 1.50(b) consists of an *R–L* load. The diode *D* is connected as shown to provide a freewheeling path for the inductor current when the BJT *Q* is turned OFF.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware breadboarding

Tasks for study:

(a) Set V_b to be a DC source. Measure the collector current and the collector–emitter voltage across transistor for different values of R_b. Tabulate the BJT collector current, base current, emitter current and collector–emitter voltage values. Find the α and β parameters. Plot the I_c–V_{ce} characteristics for different base current values.

(b) Set V_b to be a pulse source of frequency 20 kHz. Measure the collector current and the collector–emitter voltage waveforms and observe the turn-OFF and turn-ON portions.

(c) Compute the experimental switching loss and compare with the theoretical estimate.

(d) Compute the experimental conduction loss and compare with the theoretical estimate.

(e) Repeat steps (b)–(d) for different loads by varying R_L.

(f) Repeat steps (b)–(d) for different base pulse frequencies. What is the effect on the BJT power loss?

(g) What is the effect of R_b on the turn-ON and turn-OFF times? Why?

(h) Perform the steps (b)–(g) for the inductive load test circuit of Figure 1.50(b).

3. Consider the two MOSFET test circuits shown in Figure 1.51 where V_g is the gate drive pulse

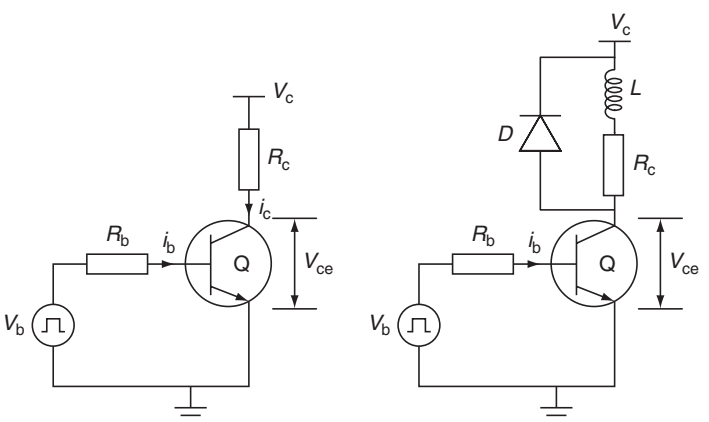


Figure 1.50 | BJT test circuit.

source. V_g should be chosen such that it is capable of both sourcing and sinking current. V_d is the drain DC supply voltage. R_d is the drain load and R_g is the gate drive resistor. The MOSFET is chosen with appropriate rating to handle the drain current and the maximum V_{ds} as decided by R_d and V_d, respectively. The circuit of Figure 1.51(a) consists of a resistive drain load and that of Figure 1.51(b) consists of an *R–L* load. The diode *D* is connected as shown to provide a freewheeling path for the inductor current when the MOSFET is turned OFF.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware breadboarding

Tasks for study:

(a) Set V_g to be a DC source. Measure the drain current and the drain–source voltage across the MOSFET for different values of V_g. Tabulate the MOSFET drain current, and drain–source voltage values. Plot the I_d–V_{ds} characteristics for different gate voltage values.

(b) Set V_g to be a pulse source of frequency 20 kHz. Measure the drain current and the drain–source voltage waveforms and observe the turn-OFF and turn-ON portions.

(c) Compute the experimental switching loss and compare with the theoretical estimate.

(d) Compute the experimental conduction loss and compare with the theoretical estimate.

(e) Repeat steps (b)–(d) for different loads by varying R_L.

(f) Repeat steps (b)–(d) for different gate pulse frequencies. What is the effect on the MOSFET power loss?

(g) What is the effect of V_g on the turn-ON and turn-OFF times?

(h) What is the effect of R_g on the turn-ON and turn-OFF times?

(i) Measure the gate current and explain its pulse shape.

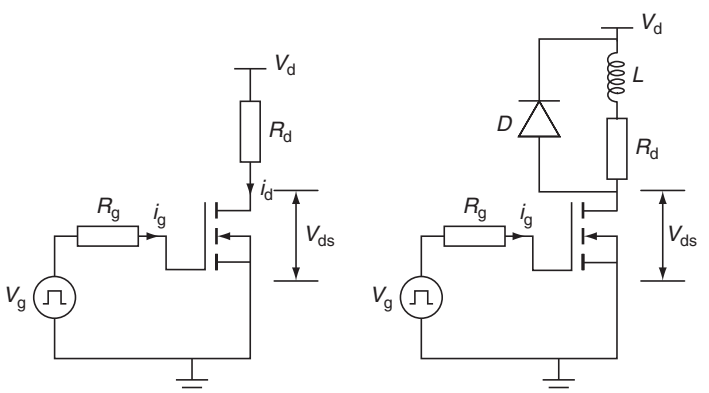


Figure 1.51 | MOSFET test circuit.

(j) Perform the steps (b)–(g) for the inductive load test circuit of Figure 1.51(b).

(k) Compare the switching speed and power loss of MOSFET with comparable BJT.

4. Consider the two IGBT test circuits shown in Figure 1.52 where V_g is the gate drive pulse source. V_g should be chosen such that it is capable of both sourcing and sinking current. V_c is the collector DC supply voltage. R_c is the collector load and R_g is the gate drive resistor. The IGBT is chosen with appropriate rating to handle the collector current and the maximum V_{ce} as decided by R_c and V_c, respectively. The circuit of Figure 1.52(a) consists of a resistive collector load and that of Figure 1.52(b) consists of an *R–L* load. The diode *D* is connected as shown to provide a freewheeling path for the inductor current when the IGBT is turned OFF.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice

b. Hardware breadboarding

Tasks for study:

(a) Set V_g to be a DC source. Measure the collector current and the collector–emitter voltage across transistor for different values of V_g. Tabulate the IGBT collector current and collector–emitter voltage values. Plot the I_c–V_{ce} characteristics for different gate voltage values.

(b) Set V_g to be a pulse source of frequency 20 kHz. Measure the collector current and the collector–emitter voltage waveforms and observe the turn-OFF and turn-ON portions.

(c) Compute the experimental switching loss and compare with the theoretical estimate.

(d) Compute the experimental conduction loss and compare with the theoretical estimate.

(e) Repeat steps (b)–(d) for different loads by varying R_L.

(f) Repeat steps (b)–(d) for different gate pulse frequencies.

(g) Perform the steps (b)–(f) for the inductive load test circuit of Figure 1.52(b).

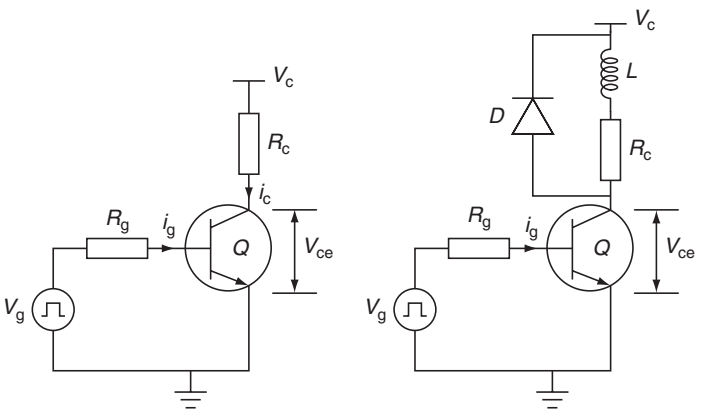


Figure 1.52 | IGBT test circuit.

FILL IN THE BLANKS

1. Diodes are devices that have two terminals: _____ and _____.
2. When the anode is more positive than the _____, the diode is said to be forward-biased.
3. When the _____ is more positive than the _____, the diode is said to be reverse-biased.
4. During turn-OFF of the diode, the _____ stored in the diffusion region has to be _____ before the junction can be reverse-biased.
5. As long as there are excess charge carriers in the diffusion region, the junction will be in _____.
6. The bipolar transistor is a _____ device.
7. BJT is a _____ device.
8. BJT is a _____ device.
9. Based on the doping type, there are two types of BJT: _____ and _____.
10. Larger the collector region, _____ is the amount of voltage that the transistor can withstand during the OFF-state.
11. The ratio of the collector current to the emitter current is called the _____ of the transistor.
12. The alpha parameter varies between _____ and _____.
13. Using the Eber-Molls model, the collector current is given as _____.
14. Using the Eber-Molls model, the emitter current is given as _____.
15. The ratio of the collector current to the base current is given by _____.
16. The electric field across the transistor is dependent on the _____ and the _____.
17. The electric field increases with _____ and with _____.
18. The safe operating areas give information about the capability of a device to handle the specified _____.
19. MOSFET is a _____.
20. The Ohmic region resistance $R_{DS(ON)}$ is due to _____.
21. $R_{DS(ON)}$ has a _____ temperature coefficient.
22. Compared to MOSFETs, BJTs have _____ conduction losses but _____ switching-losses.
23. BJTs are _____-controlled devices whereas MOSFETs are voltage-controlled devices.
24. IGBT is a _____.
25. IGBT can support _____ voltages during the OFF-state.
26. In comparison, for a given i_c, the $V_{ce(sat)}$ of the IGBT is _____ than the V_{ds} of a comparable MOSFET when ON but _____ than $V_{ce(sat)}$ of a comparable BJT.
27. IGBT has _____ overload capability as compared to a MOSFET.
28. IGBT has _____ overvoltage robustness than MOSFETs.
29. IGBTs are more like the _____ during turn-OFF and like the _____ during turn-ON.
30. Thyristors is a generic name for a semiconductor switch having _____ layers.
31. Application of a _____ voltage to the anode–cathode of an SCR will reverse-bias the SCR.
32. DIAC and TRIAC are _____ switches.
33. In a GTO, the turn-OFF is _____ whereas in an SCR, the turn-OFF is _____.

34. GTO has _____ ON-state drop as compared to the SCR.

35. The latching and holding currents are _____ for a GTO as compared to the SCR.

36. The reverse-blocking capability of a GTO is very low and hence cannot be used in _____ applications.

DESCRIPTIVE QUESTIONS

1. What are the characteristic features of an ideal switch?

2. What is a controlled switch?

3. What is an uncontrolled switch?

4. What is the depletion region?

5. *"The value of the potential barrier in volts depends on the charge carriers and is governed by the Boltzmann's relation."* Discuss.

6. With respect to the diode's *V–I* static characteristics, explain the difference among the ideal, piece-wise linear and the actual *V–I* characteristics.

7. Discuss the turn-OFF process in the diode.

8. What is meant by reverse recovery time for a diode?

9. What is the ON-state loss in a diode?

10. What is the switching loss in a diode?

11. Describe the switching process within a transistor.

12. Derive relationship for the power dissipation within the transistor between the applied collector–emitter voltage, the collector-current and the switching frequency of the transistor for (a) resistive load in the collector, (b) resistive–inductive load in the collector and c) resistive–capacitive load across the collector–emitter of the transistor.

13. Discuss forward and reverse safe operating areas.

14. What is second breakdown?

15. How are transistors connected in parallel? Discuss.

16. How does the current gain, or the beta parameter, increase with the Darlington configuration?

17. What are the semiconductor layers in a MOSFET structure?

18. What is the inversion layer? How is it formed?

19. Discuss the MOSFET structure and the various associated capacitances?

20. How are the MOSFET capacitances C_{iss}, C_{oss} and C_{rss} related to the parasitic capacitances C_{gs}, C_{gd} and C_{ds}?

21. What is gate threshold voltage?

22. What is Ohmic region of the MOSFET static characteristics?

23. Explain the MOSFET turn-OFF and turn-ON processes?

24. How is the positive temperature coefficient advantageous for paralleling of MOSFETs?

25. Discuss the functional structure of the IGBT.

26. Explain the turn-ON and turn-OFF of IGBT with waveforms.

27. What is tail current in IGBT?

28. What are holding and latching currents in SCR?

29. What is forward breakover voltage in SCR?

30. Discuss the effect of di/dt and dv/dt on the SCR switch.

PROBLEMS

1. Find the ratio of the diffusion current to the saturation current for a *p–n* junction at 75°C having a forward potential of 0.6 V applied across it.

2. A diode and a 10 Ω resistor are connected in series to a square wave voltage source of 50 V peak. Find the conduction loss for the diode if the forward dynamic resistance is (a) 0.1 Ω and (b) 0.2 Ω at the operating point. (Assume the forward barrier potential to be 0.7 V.)

3. A diode and a 10 Ω resistor are connected in series to a pulse voltage source of 50 V peak. If the forward dynamic resistance is 0.1 Ω at the operating point, then find the conduction loss of the diode when (a) pulse frequency is 20 kHz and pulse width 40 μs, (b) pulse frequency is 20 kHz and pulse width is 10 μs, (c) pulse frequency is 10 kHz and pulse width is 80 μs and (d) pulse frequency is 10 kHz and pulse width is 20 μs.

4. A diode and a 10 Ω resistor are connected in series to a square wave voltage source of 50 V peak. The reverse recovery time for the diode is given to be 200 ns. Find the switching loss of the diode when (a) input frequency is 100 kHz and (b) input frequency is 50 kHz.

5. For Problem 4 above, estimate the reverse recovery charge.

6. A BJT is driving a 10 A resistive load from a 100 V DC supply. The base drive signal is switching at frequency of 50 kHz and duty cycle of 0.75. The BJT has the following datasheet specifications: $V_{besat} = 0.7$ V, $V_{ceo} = 30$ V, $I_{cm} = 15$ A, $V_{cesat} = 0.3$ V, $h_{FEmin} = 100$, $t_d = 1$ μs, $t_s = 2$ μs, $t_r = 1.5$ μs, $t_f = 1.5$ μs. Calculate the power loss in the BJT.

7. For Problem 6 above, plot the device power dissipation as the switching frequency varies from 1 kHz to 100 kHz.

8. Two BJTs are connected in parallel to share the load current. In order to ensure sharing, two equal-valued resistors with value R are connected in the emitter leads of the BJTs. For a mismatch in the base–emitter voltage of 0.2 V between the two BJTs, a mismatch in the collector currents of the two BJTs should be less than 1 A. Calculate the value of the resistors that needs to be connected in the emitter leads.

9. A MOSFET is operated such that the operating point is in the active region. The MOSFET has a gate–source threshold voltage value of 2.5 V. A gate–source voltage of 5 V is applied to the gate–source terminals of the MOSFET which results in the flow of drain current. On increasing the gate–source voltage to 7.5 V, what is the factor by which the drain current increases?

10. A MOSFET is driving a 10 A resistive load from a 100 V DC supply. The base drive signal is switching at frequency of 100 kHz and duty cycle of 0.6. The MOSFET has the following datasheet specifications: $R_{DS(ON)} = 0.1$ Ω, $t_r = 100$ ns, $t_f = 150$ ns. Calculate the conduction and switching power losses in the MOSFET.

ANSWERS

Fill in the Blanks

1. anode and cathode
2. cathode
3. cathode; anode
4. excess charge; removed
5. forward-biased state
6. three-terminal

7. current-controlled
8. dual junction
9. NPN transistor and PNP transistor
10. greater
11. alpha parameter
12. 0.9 and 0.98
13. $I_c = \alpha_F I_F - I_{co}(e^{V_C/V_T} - 1)$
14. $I_e = \alpha_R I_R - I_{eo}(e^{V_e/V_T} - 1)$
15. β or h_{FE}
16. collector-current density; applied collector-emitter voltage
17. increasing collector voltage; increasing collector-current density
18. current and voltage
19. voltage-controlled device
20. majority carriers
21. positive
22. lower
23. current
24. voltage-controlled device
25. bipolar
26. less; greater
27. greater
28. less
29. BJTs; MOSFETs
30. four or more
31. negative
32. bi-directional
33. controlled; uncontrolled
34. higher
35. higher
36. rectifier

CHAPTER 2

Drive Circuits

Learning Objectives

After reading this chapter, you will be able to:

- design the base drive circuits for BJTs.
- design the gate drive circuits for MOSFETs and IGBTs.
- learn the principles of snubber circuits for power switches.

Almost all power applications are increasingly moving away from linear and dissipative mode of operation and towards switched-mode operation to achieve improved efficiency. The power devices now are being used primarily as switches. To achieve proper and efficient operation of the power equipments, the power devices should be driven in an appropriate manner that makes them behave as switches.

In Chapter 1, the operation, characteristics and models of the various power switches are discussed. Of the power switches discussed, the bipolar junction transistors (BJTs), the metal oxide semiconductor field effect transistors (MOSFETs) and the insulated gate bipolar transistors (IGBTs) can be switched ON and OFF by means of a control signal. This feature of controllability has made the them increasingly popular in power electronics systems like the DC–DC converters, AC-to-DC rectifiers with power factor (pf) correction, DC-to-AC inverters, DC and AC motor drives, etc. The insights gained in the previous chapter will be used here to design the drive circuits for the controlled switches. The discussion in this chapter will focus mainly on drive circuits for fully controlled power switches. Two of the generic power switch drive circuits discussed in this chapter are BJTs and MOSFETs. However it should be noted that the drive circuits for MOSFETs are directly applicable for IGBTs too. These controlled power switches account for 80–90% of the power electronic applications. In most applications MOSFETs and IGBTs are the more popular and preferred high-power switches. Their gate drives are in general composed of stages of BJT-based switches.

2.1 Transistor Drive Circuits

Figure 2.1 shows the typical collector–emitter voltage and collector-current waveforms for a switching power transistor. The turn-ON and turn-OFF intervals are indicated in Figure 2.1. The switching behavior of the transistor during these two intervals and the dependency of the turn-ON and turn-OFF behavior on the transistor base drive are discussed in the following sections.

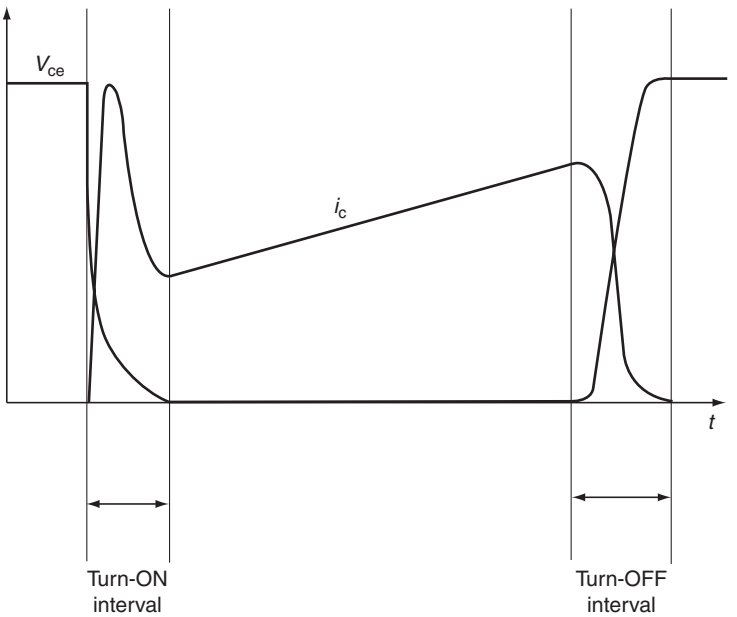


Figure 2.1 | Typical collector-current and collector–emitter voltage waveforms of a transistor.

Turn-ON Behavior

The portion of the waveform corresponding to the turn-ON of the transistor (shown in Figure 2.1) is shown in Figure 2.2 with an expanded timescale for various base-current waveshapes. A typical set of voltage and current waveforms at the collector and base of a transistor during the turn-ON interval is depicted in Figure 2.2(a). One should observe that during transistor turn-ON, a large collector-current spike, as indicated in Figure 2.2(a), is generated. Such waveforms are found in a power-converter circuit wherein a capacitance (parasitic or otherwise) is discharged at transistor turn-ON.

Figure 2.2(b) shows the turn-ON situation for a base-current waveshape that has a faster rate of rise. It can be noted that here the peak and the average values of the turn-ON dissipation are smaller than that shown in Figure 2.2(a). Figure 2.2(c) shows the effect on the transistor turn-ON for a very fast rising base-current pulse which initially overshoots the final (steady-ON) value. The turn-ON dissipation is much lower and narrower than the cases indicated in Figures 2.2(a) and 2.2(b). From Figure 2.2, it is evident that for the power transistor, the turn-ON conditions are most favorable when the driving base-current pulse has a fast leading edge and overshoots the final value or steady-ON value of the base current.

Turn-OFF Behavior

The portion of the waveform corresponding to the turn-OFF of the transistor (Figure 2.1) is shown in Figure 2.3 with an expanded timescale. For the transistor to turn-OFF faster, it is essential that a negative base-current drive be provided as indicated in Figure 2.3. The turn-OFF dissipation pulse is dependent on both the transistor turn-OFF time and the collector-current waveshape during turn-OFF. One should note that the major portion of the turn-OFF dissipation is during the fall time t_f and the dissipation during the storage time t_s is negligible.

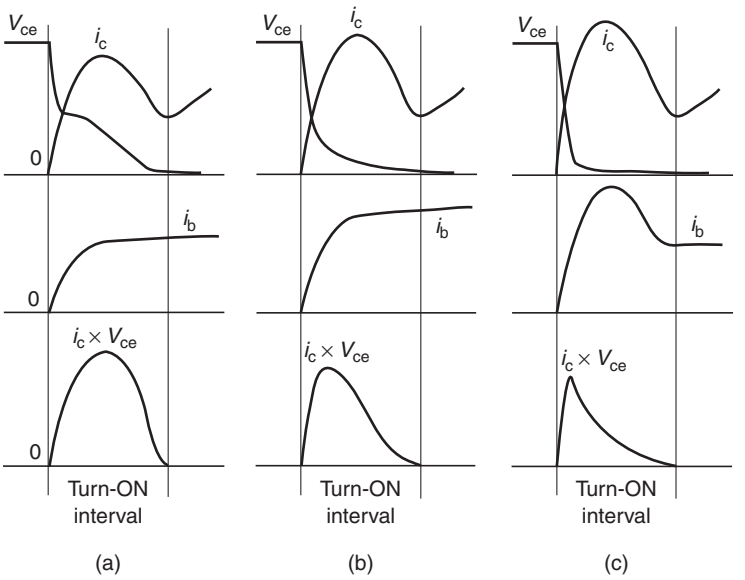


Figure 2.2 Expanded turn-ON interval to show the effect of base current on the switching characteristics.

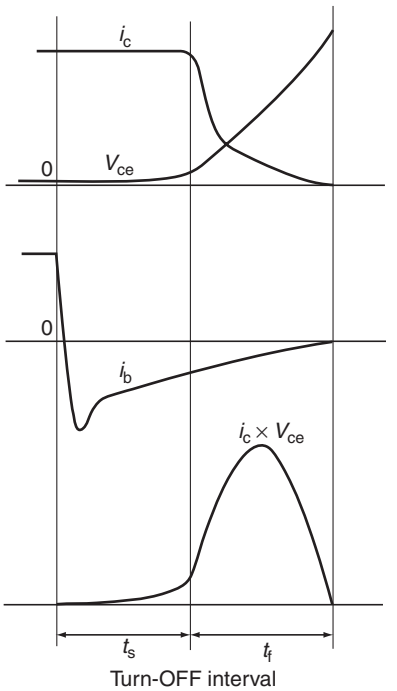


Figure 2.3 Expanded turn-OFF interval.

Characteristics and Classification of the Drive Circuits

Figure 2.4 shows the base drive requirement for a power transistor switch. During the turn-ON period, the base current should be fast rising and must overshoot the final or steady-ON value. The maximum value of the base current during turn-ON is denoted by i_{b+}. During the turn-OFF period, a negative base current (with a maximum value of i_{b-}) should be provided to quickly remove the stored charges in the transistor as explained in Chapter 1. During the steady-ON period, the base-current value is decided by the expected collector current and the minimum β ($h_{FE(sat)}$ or $h_{FE(min)}$) of the transistor. The steady-ON state base current is denoted by i_{bon}. During the steady-OFF period, the base-current value should be zero. Thus, a transistor drive circuit should be designed such that the base-current waveshape (as shown in Figure 2.4) is realized. A rule of thumb that has been found successful in most practical circuits is to take $i_{bon} = 2 \times i_c/h_{FE(sat)}$; $i_{b+} = 1.5(i_{bon})$ and $i_{b-} = 2(i_{bon})$.

A good BJT base drive circuit must be able to:

1. supply adequate positive base current during turn-ON, that is i_{b+};
2. supply adequate base current to maintain the transistor in the steady-ON state, that is i_{bon};
3. supply negative base current for proper turn-OFF of the transistor, that is, i_{b-}.

Moreover, depending on the nature of the application, the drive circuit should:

1. provide isolation between the control signals used for the base drive and the power signals that are being controlled;
2. include protection against high dv/dt or surge voltage and overcurrents.

Thus, a base drive circuit can be classified as

1. drive circuit without isolation;
2. drive circuit with transformer isolation;
3. drive circuits with opto-isolation;

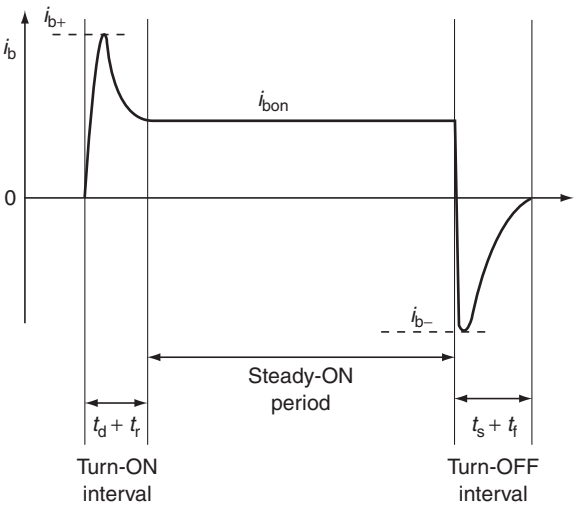


Figure 2.4 | Base drive requirement.

4. drive circuits without negative voltage source;
5. drive circuits with negative voltage source;
6. drive circuits with various combinations of the above.

Few BJT drive circuits examples are discussed in the following sub-sections.

BJT Drive Circuit-1

This base drive circuit is shown in Figure 2.5. This is the simplest of all the base drive circuits. In fact, it is the basic building block of many other complex base drive circuits. For low-power and low-frequency applications, this base drive circuit is well suited. This circuit does not provide the starting surge base current i_{b+} during the turn-ON time. The resistor R_2 provides the discharge path for removing the base charges in the transistor Q. As a rule of thumb, it is found that $i_{bon} = 2(i_c/h_{FE(sat)})$ and $i_{R2} = i_{bon}$ gives satisfactory performance. Thus

$$R_2 = \frac{V_{be(sat)}}{i_{R2}} = \frac{V_{be(sat)} h_{FE(sat)}}{2i_c} \tag{2.1}$$

$$R_1 = \frac{V_b - V_{be(sat)}}{2i_{bon}} = \frac{(V_b - V_{be(sat)}) h_{FE(sat)}}{4i_c} \tag{2.2}$$

BJT Drive Circuit-2

This drive circuit is shown in Figure 2.6. Here, Q_p is the power transistor to be controlled. Q_1 and Q_2 are auxilliary low-power transistors. This circuit is useful when the sourcing capability of V_b is low. The two transistor stages, Q_2 and Q_1, provide the necessary current gain to suitably drive Q_p. One can notice from Figure 2.6 that the drives for the intermediate transistors Q_1 and Q_2 are based on the basic BJT drive circuit-1.

When V_b is positive, Q_2 turns ON. As a consequence, collector-to-emitter voltage of Q_2 is almost zero. As R_4 is now connected to a ground at one end, the emitter–base junction of the PNP transistor Q_1 is forward-biased and therefore Q_1 turns ON. This pumps the necessary base drive current through the resistor R_1 for Q_p to turn-ON. When V_b is zero, then Q_2 is OFF and the base of Q_1 is high, thus making Q_1 OFF. This will cut off the base drive for Q_p and thus switch-OFF Q_p.

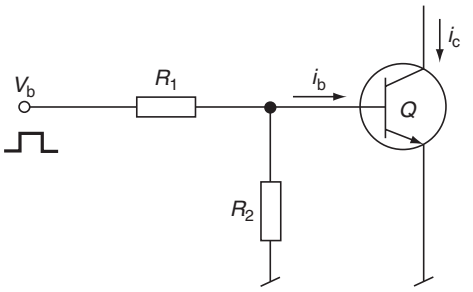


Figure 2.5 | BJT drive circuit-1.

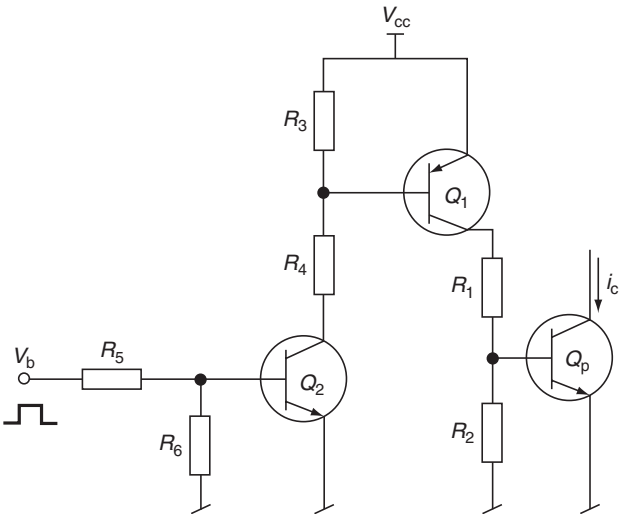


Figure 2.6 | BJT drive circuit-2.

BJT Drive Circuit-3

The base drive circuit shown in Figure 2.7 is similar to the BJT drive circuit-2 that is shown in Figure 2.6. Here, the turn-OFF time of the power transistor Q_p is improved by using an inductor L as depicted in Figure 2.7.

When the transistor Q_1 is conducting, a current is flowing through L and some energy gets stored in the inductor. The current in the inductor is limited to $i_{b\_}$ by the resistance R_2. When Q_1 is cut off, the inductor will not allow sudden change in its current, and therefore the voltage polarity across the inductor will

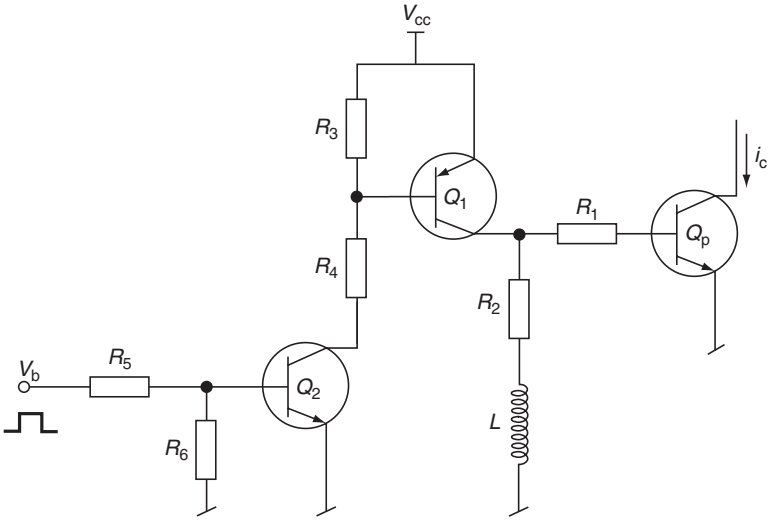


Figure 2.7 | BJT drive circuit-3.

reverse. This makes the inductor to act as a generator which will now supply the reverse base current i_{b-} for fast turn-OFF of Q_p. The value of L in μH is calculated using the following relationship:

$$L = \left| \frac{(R_1 + R_2) \cdot i_{b-} - V_{be(sat)}}{di_b / dt} \right| \tag{2.3}$$

where di_b/dt is taken as a value between $0.15i_c$ A/μs for high-voltage transistors (>700 V) to $0.5i_c$ A/μs for low-voltage transistors (<200 V).

BJT Drive Circuit-4

This base drive circuit is shown in Figure 2.8. This drive circuit is similar to the base drive circuit of Figure 2.6. But in this case, the resistor R_2 is connected to a negative supply to improve the turn-OFF speed of the power transistor Q_p. The operation is exactly similar to that of the base drive circuit-2, but here during the turn-OFF of Q_p, a negative base current i_{b-} is provided by the negative source through the resistance R_2. The resistance R_2 is decided by

$$R_2 = \frac{V_{be(sat)} + V_{cc}}{i_{b-}} = \frac{(V_{be(sat)} + V_{cc}) \cdot h_{FE(sat)}}{4i_c} \tag{2.4}$$

where i_{b-} is chosen as 2 times i_{bon}.

BJT Drive Circuit-5

This base drive circuit is shown in Figure 2.9. This drive circuit provides a fast turn-ON and turn-OFF of Q_p. This is because the circuit provides a surge base current i_{b+} during turn-ON of Q_p and also a negative base current i_{b-} during turn-OFF of Q_p.

When V_b is zero, the PNP transistor Q_3 is ON. This makes the base node of Q_1 and Q_2 positive. This will forward bias the base–emitter of Q_1 and reverse bias the emitter–base of Q_2. Thus, Q_1 is turned ON and Q_2

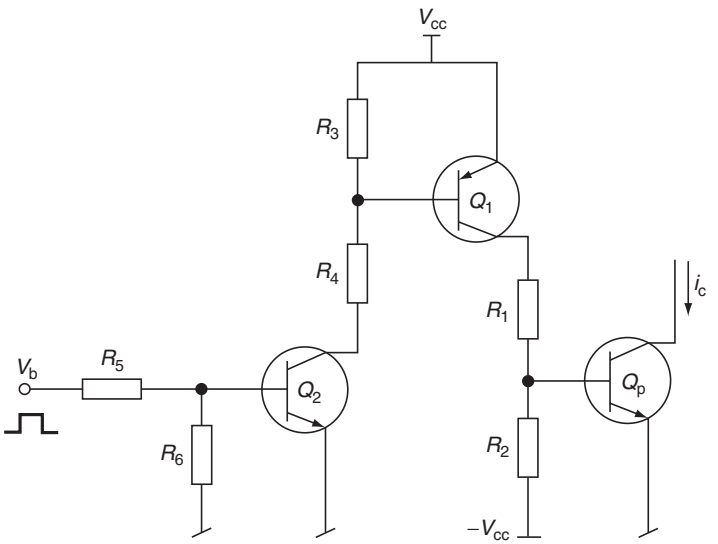


Figure 2.8 | BJT drive circuit-4.

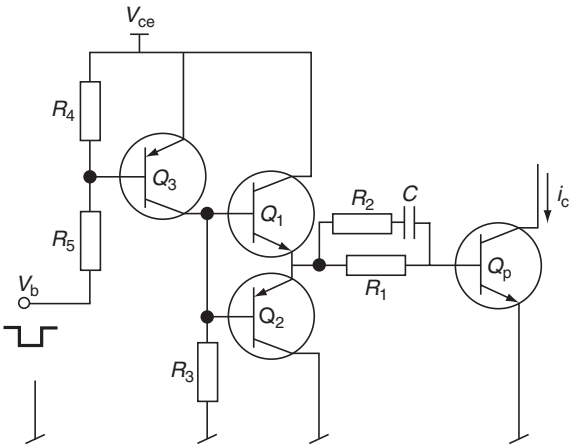


Figure 2.9 | BJT drive circuit-5.

is turned OFF. When Q_1 is turned ON, a surge current i_{b+} is delivered to the base of Q_p through the R_2–C branch. This base surge current will induce fast turn-ON of Q_p. During the time when the power transistor Q_p is fully ON, the capacitor would have charged to V_{cc}. When V_b is made positive, Q_3 is turned OFF and this makes the base node of Q_1 and Q_2 zero. The emitter of Q_2 is positive because of the voltage of V_{cc} on C. Thus, the emitter–base node of Q_2 is forward-biased and will turn ON Q_2. If emitter–base of Q_2 is forward-biased, the base–emitter of Q_1 will be reverse-biased automatically. Thus, Q_1 will be turned OFF. The capacitor C, which was charged to V_{cc}, will discharge through Q_2 providing the reverse base current i_{b-}, thus enabling fast turn-OFF of Q_p.

The turn-ON base current surge i_{b+} will flow through R_2–C branch. The steady-ON base current i_{bon} of the transistor Q_p will flow through R_1. Thus,

$$R_1 = \frac{V_{cc} - V_{ce(sat)(Q1)} - V_{be(sat)(Qp)}}{i_{bon}} \tag{2.5}$$

where $i_{bon} = 2(i_c/h_{FE(sat)})$; R_2 is chosen as 1/5 times R_1.

Now

$$\tau \approx R_2 C \tag{2.6}$$

Set τ equal to the turn-OFF time of Q_p, that is $t_{off} = t_s + t_f$. Then,

$$C = \frac{t_{off}}{R_2} \tag{2.7}$$

BJT Drive Circuit-6

This base drive circuit is shown in Figure 2.10. This drive circuit is similar in operation to the drive circuit-5 shown in Figure 2.9. Here the R_2–C branch is removed and the negative base current during turn-OFF of Q_p is provided by the negative supply as shown in Figure 2.10. In this base drive circuit, there is no turn-ON base surge current i_{b+}. However, the negative base current for fast turn-OFF is given by $(V_{cc} + V_{be(sat)})/R_1$.

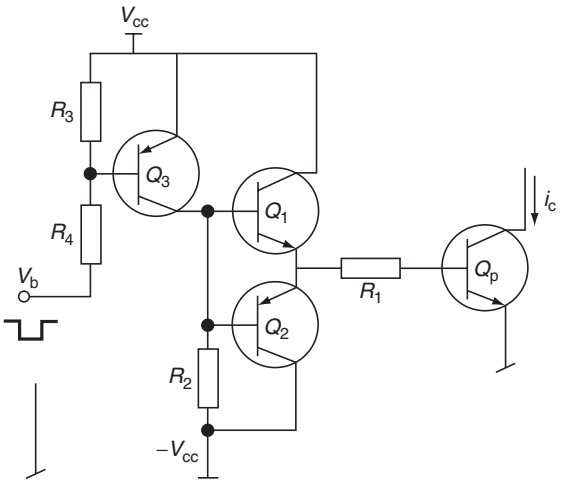


Figure 2.10 | BJT drive circuit-6.

One may recall that if a large base drive is injected, then the base charges are more and as a consequence the turn-OFF time will be longer. If, however, the transistor is allowed to be just hovering at saturation, then the stored base charge is small and the turn-OFF times will be very fast. But this will lead to higher ON-state V_{ce} of the transistor which in turn leads to larger ON-state losses. Therefore, if faster switching of the transistor is required, the transistor should be biased such that it is just out of saturation. This is done by the use of an anti-saturation circuit or the Baker's clamp as shown in Figure 2.11. Referring to Figure 2.11, it is evident that

$$V_{ce} = V_{D2} + V_{be(sat)} - V_{D1} \tag{2.8}$$

If the Baker's clamp was not used, then on providing a base overdrive, the collector–emitter voltage will reduce (to about 0.1–0.3 V). But by using the Baker's clamp, when the collector–emitter voltage reduces, diode D_1 conducts and clamps the collector–emitter voltage to approximately $V_{be(sat)}$ value (assuming approximate cancellation of drops due to D_1 and D_2). This means that V_{ce} is never lower than about 0.7 V and thus the transistor Q_p is always at the edge of saturation. Therefore, the charges in the transistor will be reduced, thereby improving the turn-OFF times. One should note that the diode D_3 is used to provide a path for the reverse base current. This anti-saturation circuit, that is, the Baker's clamp can be used in any of the base drive circuits discussed.

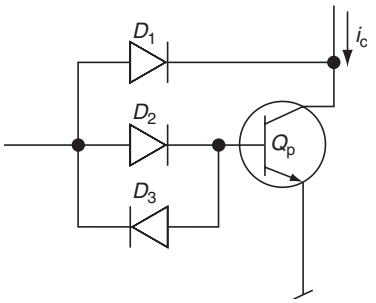


Figure 2.11 | Anti-saturation circuit or Baker's clamp.

BJT Drive Circuit-7

This base drive circuit is shown in Figure 2.12. The base drives discussed until now are for a particular value of collector currents. If the collector current (i.e., load) decreases, the transistor will be overdriven into saturation for the same base drive. On the other hand, if the collector current increases, the transistor will come out of saturation and operate in the active region. In applications where the load current can vary within a wide range, the above problem can be overcome if the base current also varies in proportion to the collector current. This concept is called the proportional-base drive. The base drive circuit indicated in Figure 2.12 is a proportional-base drive circuit.

A current transformer (CT) is placed in the collector as shown in Figure 2.12. A short-duration positive pulse (which is generally obtained by passing the base drive pulse through an *R–C* differentiator network) is applied to the base of Q_p. The energy in the base drive pulse is just sufficient to turn-ON the transistor Q_p. As a result, the collector current increases. Owing to transformer action, the collector current gets reflected into the secondary of the CT and a current i_{n2} will flow through the base of Q_p. This increased drive will further turn ON Q_p which will, in turn, increase i_c. In this manner, the transistor will go into saturation by re-generative action. The turns ratio for the CT is chosen in the following manner:

$$\frac{i_b}{i_c} = \frac{1}{h_{FE(min)}} = \frac{i_{n2}}{i_c} = \frac{n_1}{n_2} \tag{2.9}$$

One can note from Eq. (2.9) that the base current to the transistor Q_p varies in proportion to the load current, that is, the collector current. Therefore, the turn-ON and turn-OFF speeds will be fast at all loads.

To turn OFF the transistor, a negative pulse is applied as shown in Figure 2.12. This will bring the transistor out of saturation and will cause the collector current to decrease. This will lead to re-generative turn-OFF of the transistor. The transistor Q_1 is provided to demagnetize the core of the CT. During the time when Q_p is OFF, Q_1 is turned ON. This causes a current to flow through the winding n_3 in a direction which will re-set the core of the CT. The design of the CT is discussed in Chapter 7.

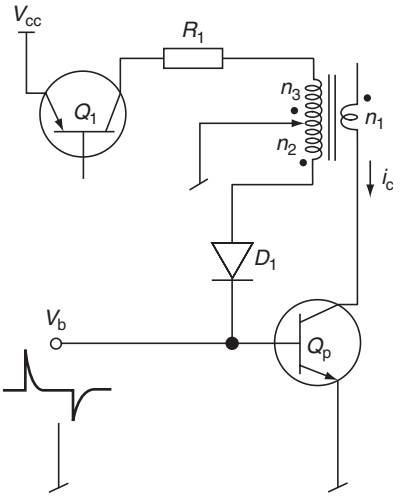


Figure 2.12 | BJT drive circuit-7.

BJT Drive Circuit-8

Until now, all the base drive circuits considered had no isolation between the base or control portion of the circuitry and the load or the power portion of the circuitry. However, in many situations isolation between the high-power side and the low-power base side of the circuit is essential. In such cases, the base drive circuits will have to be designed taking isolation into account. Generally, isolation can be provided either by using opto-couplers or by using transformers. The base drive circuit shown in Figure 2.13 is one such isolated drive circuit which uses an opto-coupler to provide the isolation between the base side and the high-power collector side circuits.

One can note that the base drive circuit portion in Figure 2.13 after the opto-isolation is exactly similar to the base drive circuit-5 shown in Figure 2.9. Therefore, the turn-ON and turn-OFF operations of the power transistor Q_p are similar. A positive base drive pulse V_b causes a current to flow through the photodiode of the opto-coupler. This causes the transistor Q_4 of the opto-coupler to turn ON. As a consequence, Q_3 is OFF and Q_1 is ON as discussed for the base drive circuit-5 of Figure 2.9. Likewise, when V_b is zero, no current flows through the photodiode and Q_4 is OFF. This causes Q_3 to turn ON. As a consequence, the power transistor Q_p turns OFF with reverse base current flowing through Q_2. One should note that the choice of the opto-coupler depends on the current transfer ratio of the opto-coupler, the propagation delay for turn-ON and turn-OFF and the breakdown voltage between the opto-coupler diode and transistor terminals.

BJT Drive Circuit-9

This is a transformer-isolated base drive circuit and is shown in Figure 2.14. When a positive voltage V_b is applied to Q_1, the transistor Q_1 will turn ON and the full voltage of V_{cc} is applied across the primary of the transformer. The transformer is a pulse transformer designed for the specific frequency and it passes the primary pulse to the secondary which turns ON Q_p. When the transistor Q_1 is turned OFF, the current in the primary will be suddenly cut off and will give rise to a large voltage (of $L\mathrm{d}i/\mathrm{d}t$) across the primary. This will also appear across Q_1 (by Kirchhoff's voltage law) and will damage the transistor. Therefore, to prevent the abrupt cut off of the primary current when Q_1 turns OFF, a freewheeling path through D_f and R_3 is provided. (The design of the transformer is discussed in Chapter 7.)

The ratio of the time for which the power transistor Q_p is ON to the switching time period is called the *duty cycle or the duty ratio, D*. The base drive circuits 1–8 discussed until now can be used for duty cycle range

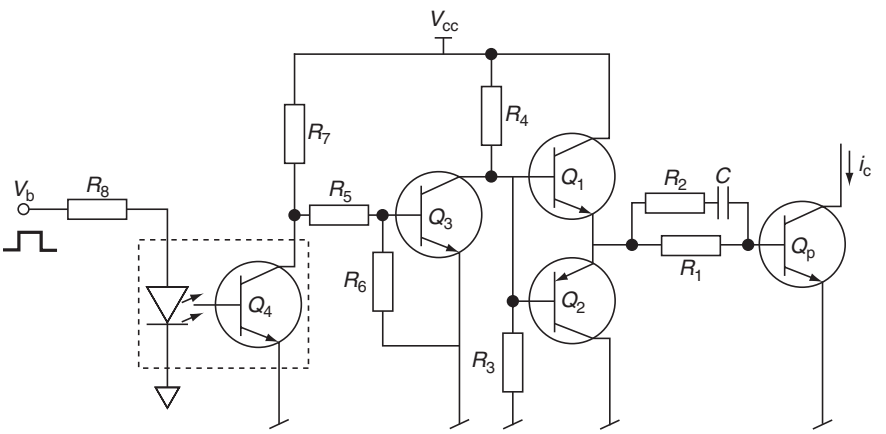


Figure 2.13 | BJT drive circuit-8.

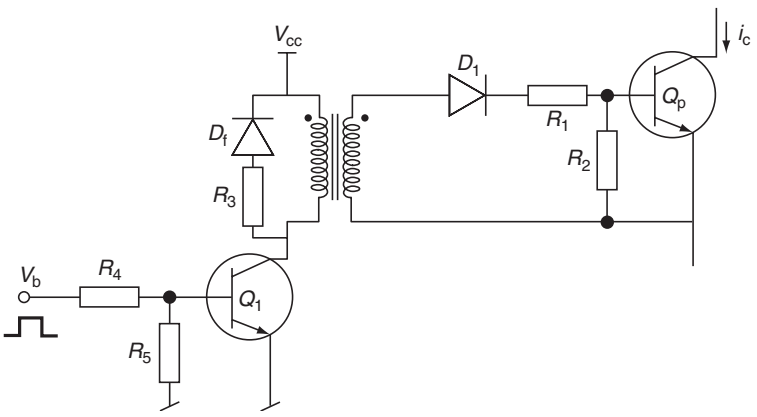


Figure 2.14 | BJT drive circuit-9.

from 0 to 1. But in this case, the duty cycle is limited because the transistor has to be OFF for the period of time when the core is being re-set by freewheeling action. If the value of R_3 is large, then the core re-setting is faster and therefore the range of duty cycle is larger as Q_p needs to be compulsorily OFF only for a smaller time. However, in such a case, as the drop across R_3 is higher, the V_{ceo} rating of Q_1 is higher which is evident from

$$V_{ceo(Q1)} > V_{cc} + I_{mag} R_3 + V_{Df} \tag{2.10}$$

where I_{mag} is the magnetizing current in the primary at the instant when Q_1 is being turned OFF. Therefore, to have a reasonable V_{ceo} rating for Q_1, the value of R_3 cannot be chosen too large. In practice, the duty ratio is limited to less than 0.5 (50%).

BJT Drive Circuit-10

This base drive circuit is shown in Figure 2.15. This is exactly similar to the base drive circuit of Figure 2.14. The only difference lies in the freewheeling circuitry. In Figure 2.14, the freewheeling path contained

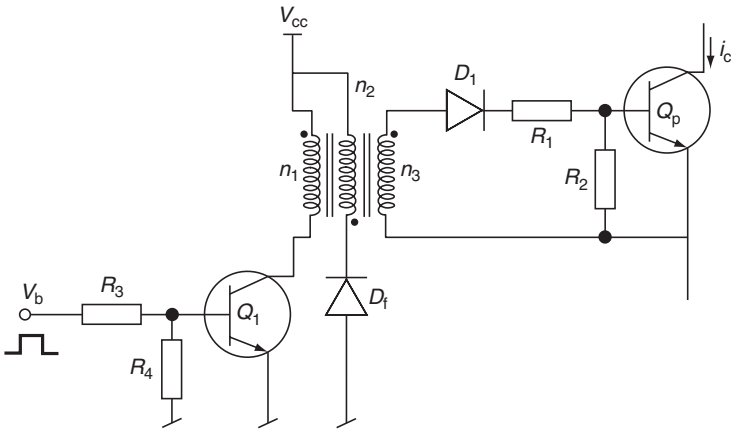


Figure 2.15 | BJT drive circuit-10.

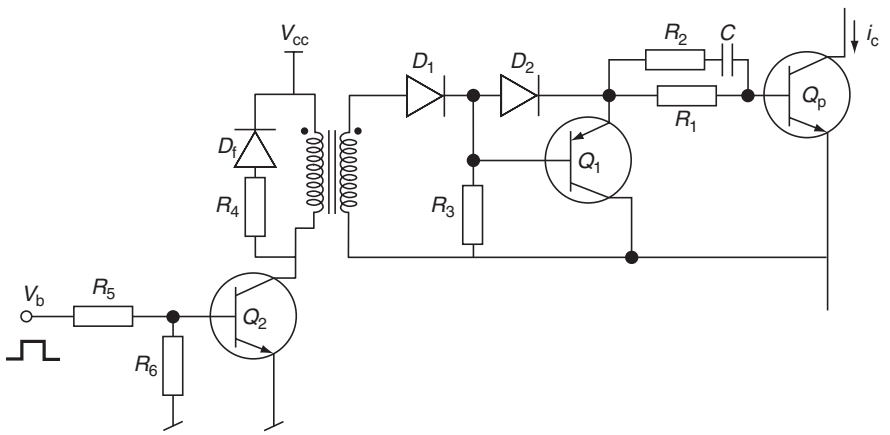


Figure 2.16 | BJT drive circuit-11.

a resistor R_3 which leads to dissipation. In Figure 2.15, the dissipation is avoided during freewheeling. When Q_1 turns OFF, the dot poles of the transformer become negative with respect to the other pole. This causes D_f to be forward baised. D_f conducts and freewheels the magnetic energy stored in the core. This causes the required core re-setting which prevents core saturation. As winding n_2 is used for demagnetizing the core, this winding is also called the demagnetizing winding. In practice, to achieve a very tight coupling (i.e., low leakage) between n_1 and n_2, these two windings are wound bifilar. Therefore, the turns ratio n_1:n_2::1:1 is maintained. As a consequence, a time equal to the ON time of Q_1 is required for the core to re-set. Therefore, the duty cycle in this case cannot exceed 0.5 (or 50%).

BJT Drive Circuit-11

This is a transformer-isolated base drive circuit as shown in Figure 2.16. This circuit is similar to the transformer-isolated base drive circuits of Figures 2.14 and 2.15. But here, the turn-OFF time of Q_p is improved by using an active turn-OFF circuit with the help of a PNP transistor Q_1.

When V_b is positive, the transformer secondary is positive at the dot pole. This will forward bias the diodes D_1 and D_2. The drop across D_2 will ensure that the emitter–base junction of Q_1 is reverse-biased. Thus, Q_1 is in the turned-OFF condition. The R_2–C will provide a low-impedance path during pulse transitions which will provide the needed surge base current for fast turn-ON, that is i_{b+}. When Q_p is to be turned OFF, V_b is made zero. Then, D_1 will be reverse-biased and the base of Q_1 is pulled low. Owing to the presence of the capacitor C, the emitter of Q_1 will now be positive which ensures that the emitter–base junction of Q_1 is forward-biased. This causes Q_1 to turn ON, thereby providing the necessary reverse base current for fast turn-OFF of Q_p. One should note that here also the duty cycle is limited to 0.5.

BJT Drive Circuit-12

This base drive circuit is shown in Figure 2.17. This is an isolated proportional-base drive circuit. As discussed in the case of the non-isolated proportional-base drive circuit of Figure 2.12, there is a CT (n_c – n_b windings) which provides the necessary base current to maintain the power transistor Q_p in the ON-state. A flyback tranformer with air gap (n_p – n_b windings) is used to supply the required turn-ON and turn-OFF pulses to start the turn-ON and turn-OFF re-generative processes, respectively, for Q_p.

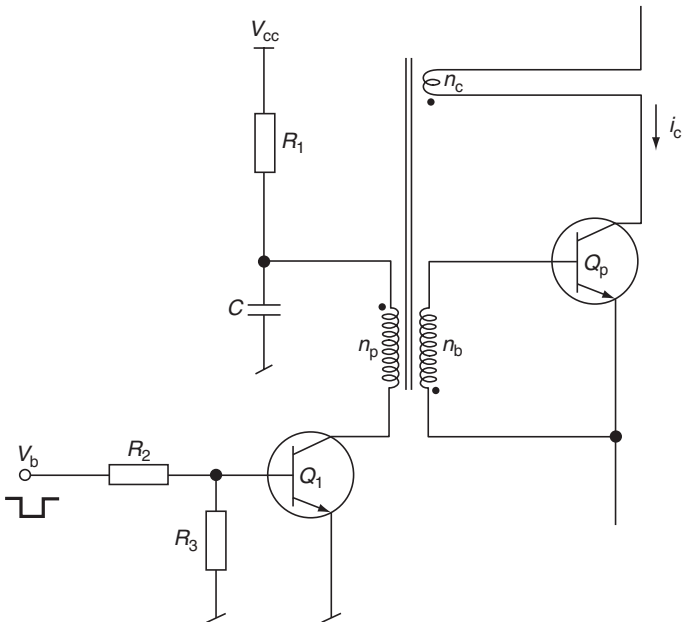


Figure 2.17 | BJT drive circuit-12.

The operation of the base drive circuit of Figure 2.17 is similar to the non-isolated proportional-base drive circuit. Consider the situation where Q_p is in the ON condition. During this time, Q_1 is OFF. The capacitor C charges to V_{cc} through R_1 with a time constant of R_1C. Now, when Q_1 is turned ON, the secondary voltage will be the same as the base–emitter voltage of Q_p which is around 0.7 V because as yet Q_p is not turned OFF due to the presence of stored charges in it. Therefore, the primary of the transformer will experience a virtual short circuit. A large current will be discharged from C. This will be reflected at the secondary as a large negative base current which will quickly and re-generatively turn OFF Q_p.

During the time when Q_p is OFF and Q_1 is ON, the primary current will rise and saturate the core. The resistance R_1 will limit the current through Q_1. The dot poles of the transformer are positive and therefore Q_p is maintained in the OFF-state. Now, if Q_1 is turned OFF, the voltage polarities across the windings reverse due to inductance action and the stored energy in the core (air gap) will freewheel through the base of Q_p. This will provide the turn-ON energy to start the re-generative process for turn-ON of Q_p.

The choice of n_c:n_b will depend on the $h_{FE(min)}$ of the power transistor Q_p. Thus

$$\frac{n_b}{n_c} = \frac{i_c}{i_b} = h_{FE(min)} \tag{2.11}$$

The choice of n_p:n_b will depend on the base–emitter breakdown voltage of Q_p. This is because, when Q_p is OFF, the secondary voltage should not exceed the base–emitter breakdown voltage BV_{ebo} of Q_p. Thus,

$$\frac{n_p}{n_b} = \frac{V_{cc}}{BV_{ebo}} \tag{2.12}$$

When Q_1 is ON, the current through the primary increases till it saturates the core. This saturation current i_{pon}, which is limited by R_1, should provide the necessary turn-ON base drive current i_{b+} for Q_p. Thus,

$$i_{pon} = i_{b+}\left(\frac{n_b}{n_p}\right) \tag{2.13}$$

$$R_1 = \frac{V_{cc}}{i_{pon}} \tag{2.14}$$

When Q_1 is turned ON in order to turn OFF Q_p, the primary current i_{poff}, consists mainly of the reflected collector current of Q_p and the reverse base drive current i_{b-} of Q_p. Thus,

$$i_{poff} = i_c\left(\frac{n_c}{n_p}\right) + i_{b-}\left(\frac{n_b}{n_p}\right) \tag{2.15}$$

This current of i_{poff} has to be supplied by the capacitor when Q_1 is turned ON. Thus, the capacitor should have energy of at least

$$E_c = \frac{1}{2}CV_{cc}^2 = V_{cc} \cdot i_{poff} \cdot t_{off} \tag{2.16}$$

where t_{off} is the turn-OFF time of Q_p which is equal to $(t_s + t_f)$ of Q_p. From Eq. (2.16), the value of capacitor C is given by

$$C = \frac{2 \cdot i_{poff} \cdot t_{off}}{V_{cc}} \tag{2.17}$$

One should note that when Q_p is ON, the capacitor should charge to V_{cc} during this time. The charging time constant is R_1C. The capacitor will charge fully to V_{cc} in 5 R_1C. This means that Q_p should be on for at least 5 R_1C. Therefore, there is a minimum duty cycle limitation in this circuit. Thus, the duty cycle ranges from $5R_1C/T_s$ to almost 1, where T_s is the switching period.

BJT Drive Circuit-13

This drive circuit is shown in Figure 2.18. This is almost identical to that shown in Figure 2.17. The only difference is that the charging time constant of C is made faster here by using a transistor Q_2. It can be shown that the capacitor charging time constant is given by

$$\tau = \frac{R_1C}{h_{FE(Q2)}} \tag{2.18}$$

It is evident from Eq. (2.18) that the charging time constant is significantly reduced if Q_2 is a high h_{FE} transistor. The h_{FE} of Q_2 is generally chosen such that τ is half of the turn-ON time of Q_p that is $(t_r + t_d)$. With this circuit, the duty cycle range is increased from almost 0 to almost 1.

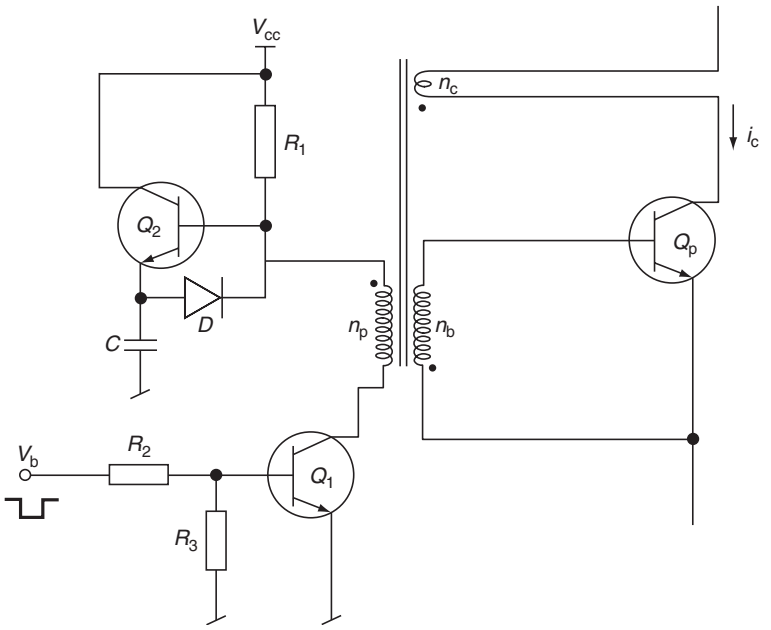


Figure 2.18 | BJT drive circuit-13.

2.2 MOSFET Drive Circuits

The MOSFETs are voltage-controlled devices. As a consequence, the gate currents are not dependent on the drain currents. The gate power required to maintain the MOSFETs in the ON condition is negligible. This section discusses primarily the MOSFET drive circuits; however, as discussed in Chapter 1, the gate portion of the IGBTs being functionally similar to the MOSFETs, these drive circuits that will be discussed can also be applied for driving IGBTs.

Figure 2.19 shows the gate drive requirements for a MOSFET. It can be observed from the gate drive requirements that the gate current required to maintain the MOSFET in the steady-ON condition is zero. Therefore, the gate power required to maintain the MOSFET in the steady-ON condition is low. The gate circuit energy is used only to turn-ON and turn-OFF the MOSFET. During turn-ON, a peak current of i_{g+} is applied which is used to turn-ON the MOSFET. During turn-OFF, a negative peak current of i_{g-} is provided for fast turn-OFF of the MOSFET.

During the turn-ON period (t_{on}), it can be assumed that an equivalent constant current of i_{gon} is being applied to the gate of the MOSFET. One can approximate the turn-ON surge current to be a right-angled triangle with the peak of i_{g+}. Thus

$$i_{gon} = \frac{i_{g+}}{2} \tag{2.19}$$

The turn-ON and turn-OFF operation of the MOSFET is as explained in Chapter 1. To turn-ON a MOSFET, a specific amount of gate charge Q_G has to be supplied to the gate of the MOSFET. This amount of Q_G for a specific MOSFET is given in the manufacturers' datasheets. Thus

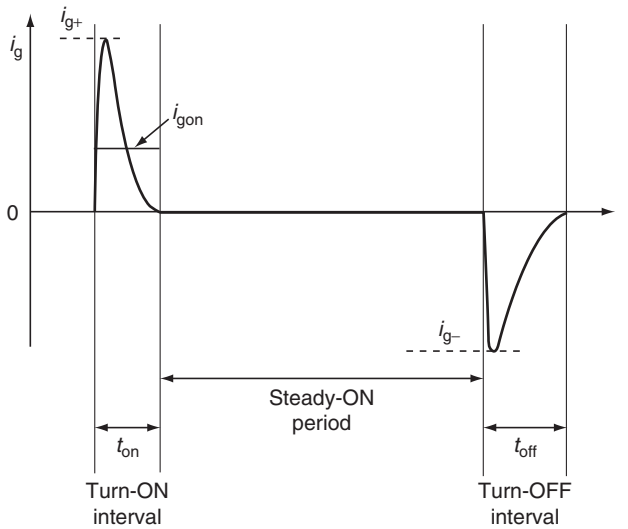


Figure 2.19 | Gate drive requirement for MOSFETs.

$$Q_G = i_{gon} t_{on} \tag{2.20}$$

It is evident from Eq. (2.20) that if one requires to turn-ON the MOSFET faster (i.e., smaller t_{on}), the i_{gon} required should be more. If the MOSFET can be switched slower, then a smaller i_{gon} would suffice. Consider a MOSFET where 250 nC of charge is required to turn-ON the MOSFET. For a specific application, if it is required that the MOSFET should be turned-ON in 1 μs, then the i_{gon} required would be 250 mA and i_{g+} required would be 500 mA. If on the other hand, the required turn-ON time is 2 μs, then i_{gon} required would be 125 mA and i_{g+} required would be 250 mA.

To allow a turn-ON gate drive surge current of i_{g+}, only a limiting resistor R_1 can be connected in series with the gate of the MOSFET such that

$$R_1 = \frac{V_{cc}}{i_{g+}} \tag{2.21}$$

where V_{cc} is the gate drive supply voltage.

MOSFETs generally require a gate voltage of 15 V. Therefore, V_{cc} = 15 V in most cases. Various MOSFET gate drive circuits will be now discussed in the following sections.

MOSFET Drive Circuit-1

The MOSFET drive circuit shown in Figure 2.20 is the simplest MOSFET drive circuit. The resistor R_1 is used for limiting the turn-ON surge current and it is chosen based on Eq. (2.21). If the source V_g does not have current-sinking capability, a resistor R_2 is connected as shown and is used to provide a discharge path for the input capacitance during turn-OFF of the MOSFET. R_2 is usually chosen 10 times the value of R_1.

If the gate drive source V_g is capable of sinking current, then R_2 need not be used. In such a case, as the output resistance of the source V_g is low, the input capacitance will charge and discharge through R_1.

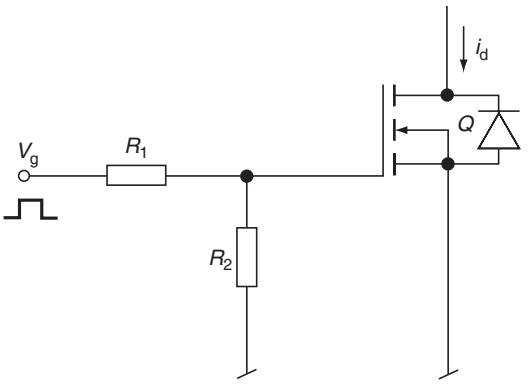


Figure 2.20 | MOSFET drive circuit-1.

MOSFET Drive Circuit-2

MOSFETs can be driven directly from a complementary metal oxide semiconductor (CMOS) logic integrated circuit (IC) as shown in Figure 2.21. For faster switching, i_{gon} should be increased. This is achieved by paralleling buffers as indicated in Figure 2.21. Either the CMOS hex inverting buffer IC or non-inverting hex buffer IC can be used to drive the MOSFETs. The input gate pulse signal source, V_g should have a polarity that should be appropriate with respect to the type of buffer chosen, whether inverting or non-inverting. It should be noted that here also as in all the MOSFET gate drive circuits that are discussed, the series resistor R_1, which is used to limit the turn-ON gate surge current, is selected based on Eq. (2.21).

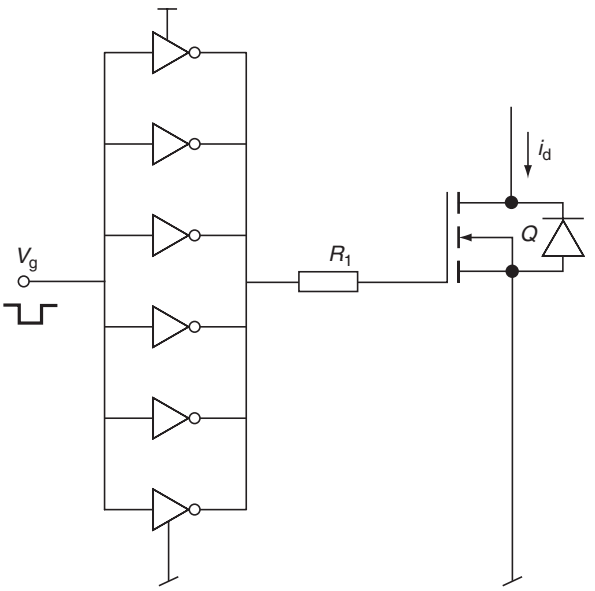


Figure 2.21 | MOSFET drive circuit-2.

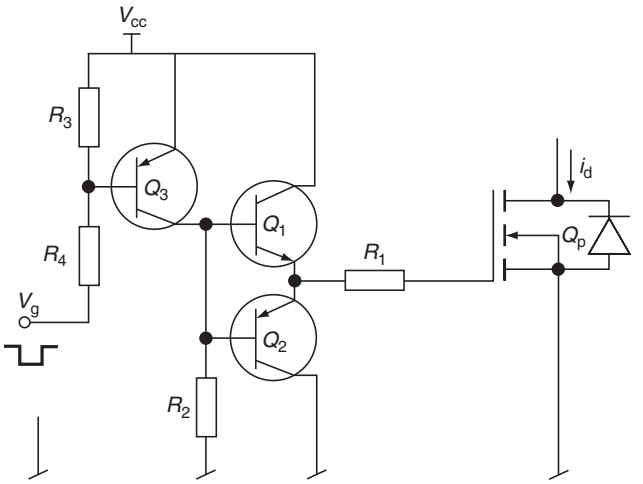


Figure 2.22 | MOSFET drive circuit-3.

MOSFET Drive Circuit-3

The MOSFET drive can be made faster by using transistors to increase the turn-ON gate drive current, i_{gon}. The MOSFET drive circuit shown in Figure 2.22 is similar in topology and operation to the BJT drive circuit shown in Figure 2.10. The only difference is that the negative power supply is not used here. In Figure 2.10, the negative power supply was used to provide the forward bias for Q_2 while turning OFF Q_p. Here, the input capacitance of the MOSFET provides the necessary forward bias to turn ON the PNP transistor Q_2 when Q_1 is turned OFF in order to switch-OFF the MOSFET Q_p. This would imply that a negative power supply is not needed.

However, if a lower voltage negative power supply V_{ss} (~–5V) is used wherein the emitter of Q_2 is connected to the negative pole of the V_{ss}, then Q_2 would be ON even when the gate–source MOSFET capacitor is fully discharged. This would ensure that the gate–source is connected by R_1 which is almost like a short as R_1 is a small resistance value. This would also ensure greater reliability in the sense that no spurious dv/dt on the high-power drain side circuit will charge up the gate–source capacitance through the drain–gate capacitance due to the fact that the gate is connected to source through R_1 and Q_2.

MOSFET Drive Circuit-4

Some circuits demand that the gate and source terminals of MOSFETs are floating with respect to those of other MOSFETs in the circuit. This is typical of bridge topologies wherein the top MOSFET and the bottom MOSFET need different drive grounds. Isolated drive to these MOSFETs can be provided either by using opto-isolators or by using pulse transformers. Figure 2.23 shows a gate drive circuit using an opto-isolator. This circuit is similar in topology and operation to the BJT opto-isolated circuit shown in Figure 2.13. One should note that for all opto-isolated drive circuits whether BJT or MOSFET drive circuits, the secondary-side circuitry (transistor side) should have a local power supply which is isolated from the primary-side circuitry (diode side).

When V_g is positive, the diode in the primary side of the opto-isolator emits photons. These photons impinge on the base region of the transistor in the secondary side. This causes photogeneration of carriers

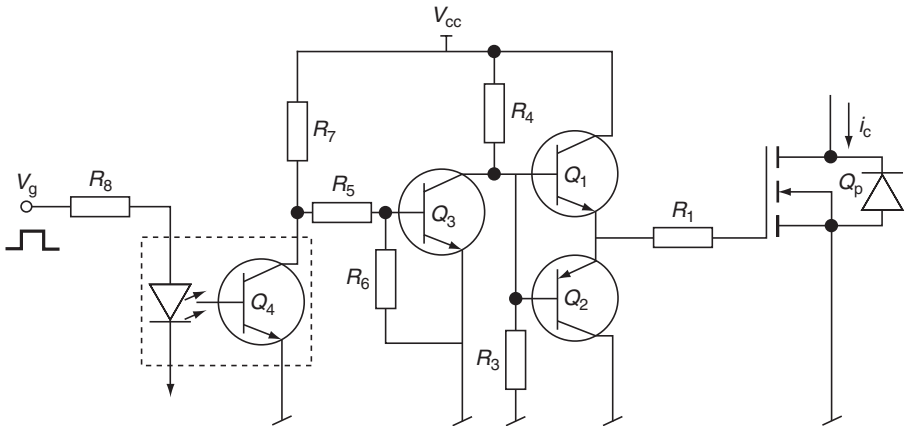


Figure 2.23 | MOSFET drive circuit-4.

sufficient to satisfy the base requirement for turn-ON of Q_4. When Q_4 turns ON, Q_3 is turned OFF. This causes Q_1 to turn ON which supplies the required gate current to the MOSFET through R_1 and turns ON the MOSFET Q_p.

MOSFET Drive Circuit-5

The gate drive circuit shown in Figure 2.24 is another opto-isolated circuit. Here, the complementary pair Q_1 and Q_2 shown in Figure 2.23 is replaced by CMOS buffer ICs either inverting or non-inverting type as required by the polarity of V_g.

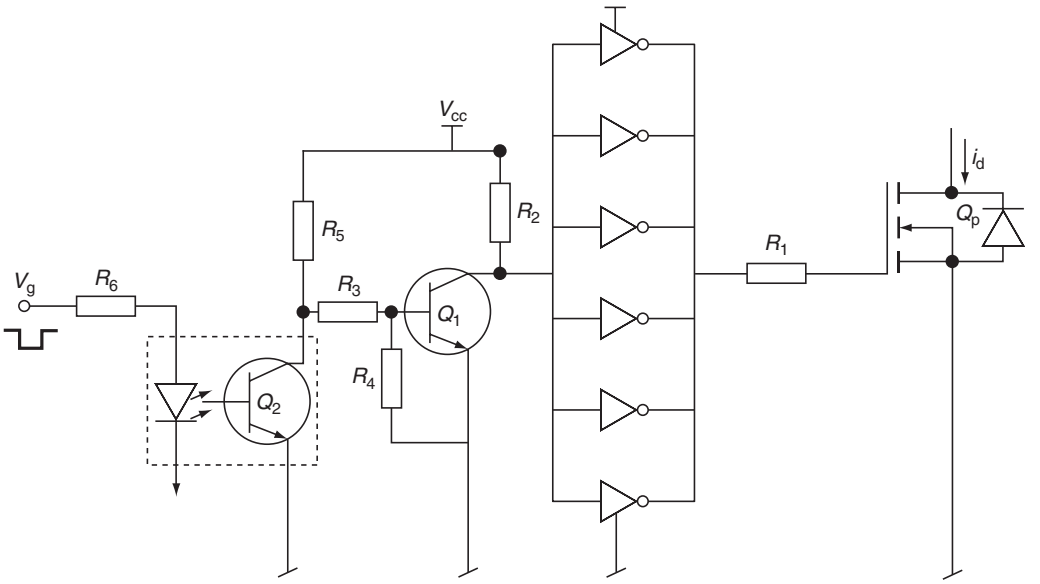


Figure 2.24 | MOSFET drive circuit-5.

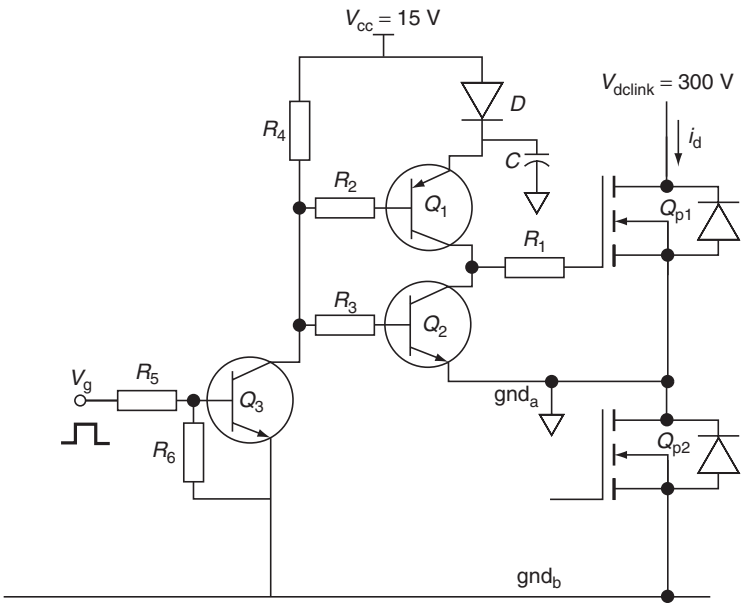


Figure 2.25 | MOSFET drive circuit-6.

MOSFET Drive Circuit-6

In most inverter and converter applications, the power converter is generally a half-bridge or a full-bridge configuration. The top MOSFET and the bottom MOSFET are switched at mutual exclusive time intervals to prevent shorting the DC link or bus. In either case, the gate drive for the high-side MOSFET is generally isolated from the low-side MOSFET by the use of opto-isolators or pulse transformers. Figure 2.25 gives a circuit wherein the high-side and the low-side MOSFETs can be driven without the use of opto- or transformer isolation.

When V_g is zero, both Q_3 and Q_1 are OFF. Therefore, the MOSFET Q_{p1} is also in the OFF-state. But during this period, the bottom MOSFET Q_{p2} is ON. The capacitor C will get charged to V_{cc} through the diode D and the bottom ON MOSFET Q_{p2}. When Q_{p2} is turned OFF and V_g is positive, Q_3 is ON, which in turn switches Q_1 ON. The capacitor C now discharges through Q_1 into the input capacitance of the MOSFET Q_{p1} thereby turning ON Q_{p1}. When V_g is made zero again, Q_1 is turned OFF and Q_2 is turned ON. The input capacitor of the MOSFET discharges through R_1 and Q_2, thereby turning Q_{p1} OFF.

As the capacitor C needs to charge to V_{cc} during the time when Q_{p2} is ON, there is a minimum ON-time limitation on Q_{p2}. It should also be noted that when Q_{p1} is ON and Q_{p2} is OFF, the Q_{p1} drive capacitor's negative pole which is connected to the source of Q_{p1} is at the DC-link voltage. At this time the diode D will have a reverse voltage of $(V_{cc} - V_C - V_{dclink})$ and therefore should be rated to withstand the full DC-link voltage.

MOSFET Drive Circuit-7

A transformer-isolated gate drive circuit is shown in Figure 2.26. One can note that this gate drive circuit is exactly similar both in topology and operation to the BJT transformer-isolated base drive circuit shown in Figure 2.14. The operation of the pulse transformer and the freewheeling circuit is similar to that discussed for the transformer-isolated BJT base drive circuit. Here again, one can use a demagnetizing winding to perform non-dissipative freewheeling as indicated in Figure 2.15 for the BJT base drive.

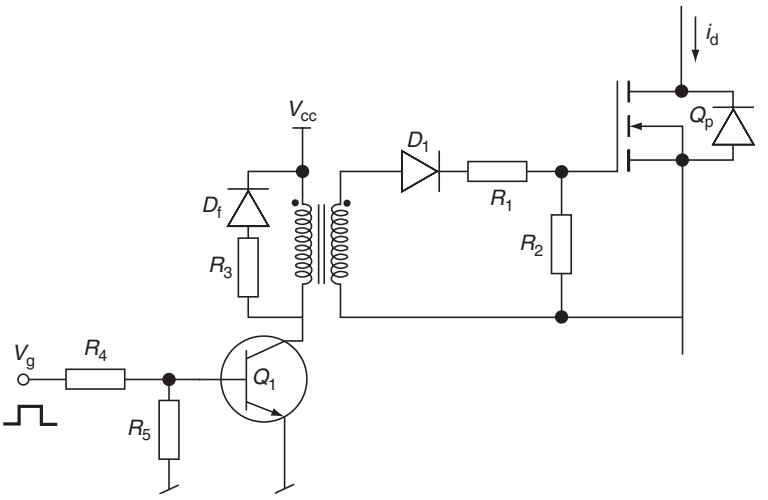


Figure 2.26 | MOSFET drive circuit-7.

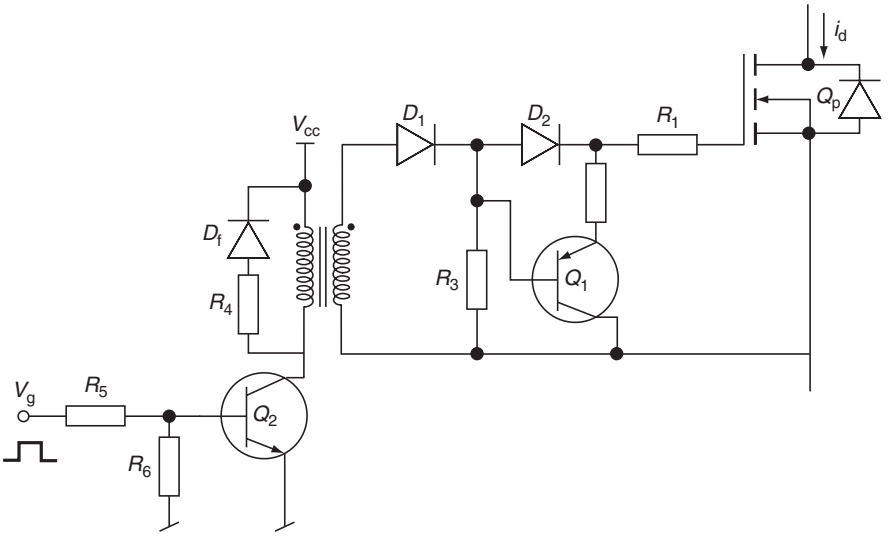


Figure 2.27 | MOSFET drive circuit-8.

For the MOSFET gate drive circuits discussed till now, the upper limit on the duty cycle is 1 (i.e., 100%). But in the case of this drive circuit, the upper limit on the duty cycle is 0.5, that is, 50% only.

MOSFET Drive Circuit-8

Another transformer-isolated gate drive circuit is shown in Figure 2.27. Here, the turn-OFF is due to active pull down by the PNP transistor Q_1. This will, therefore, provide a faster turn-OFF of the MOSFET than

that provided by the gate drive circuit shown in Figure 2.26. When V_g is positive, Q_2 is turned ON. This causes the dot poles of the pulse transformer to be positive with respect to the other poles. D_1 and D_2 will be forward-biased and charge the input capacitor of the MOSFET and turn-ON Q_p. During this time, as D_2 is ON, the emitter–base junction of Q_1 is reverse-biased and therefore Q_1 is OFF. When V_g is made zero, Q_2 turns OFF. This causes the dot poles of the pulse transformer to become negative with respect to the other poles. This will reverse bias D_1 and D_2, thereby switching OFF these diodes. As a consequence, the base of Q_1 is pulled low through R_3. The charge on the input capacitance of the MOSFET will make the emitter of Q_1 positive with respect to the base and will therefore forward bias the emitter–base junction of Q_1 and turn it ON. The input capacitance will then discharge through Q_1, thereby turning Q_p OFF. One should note that here also the upper limit for the duty cycle is 0.5, that is, 50%.

MOSFET Drive Circuit-9

To obtain the full duty cycle range of 0 to 1 using transformer isolation, one can use the transformer-isolated gate drive circuit shown in Figure 2.28. Here, two pulse transformers are used and they are

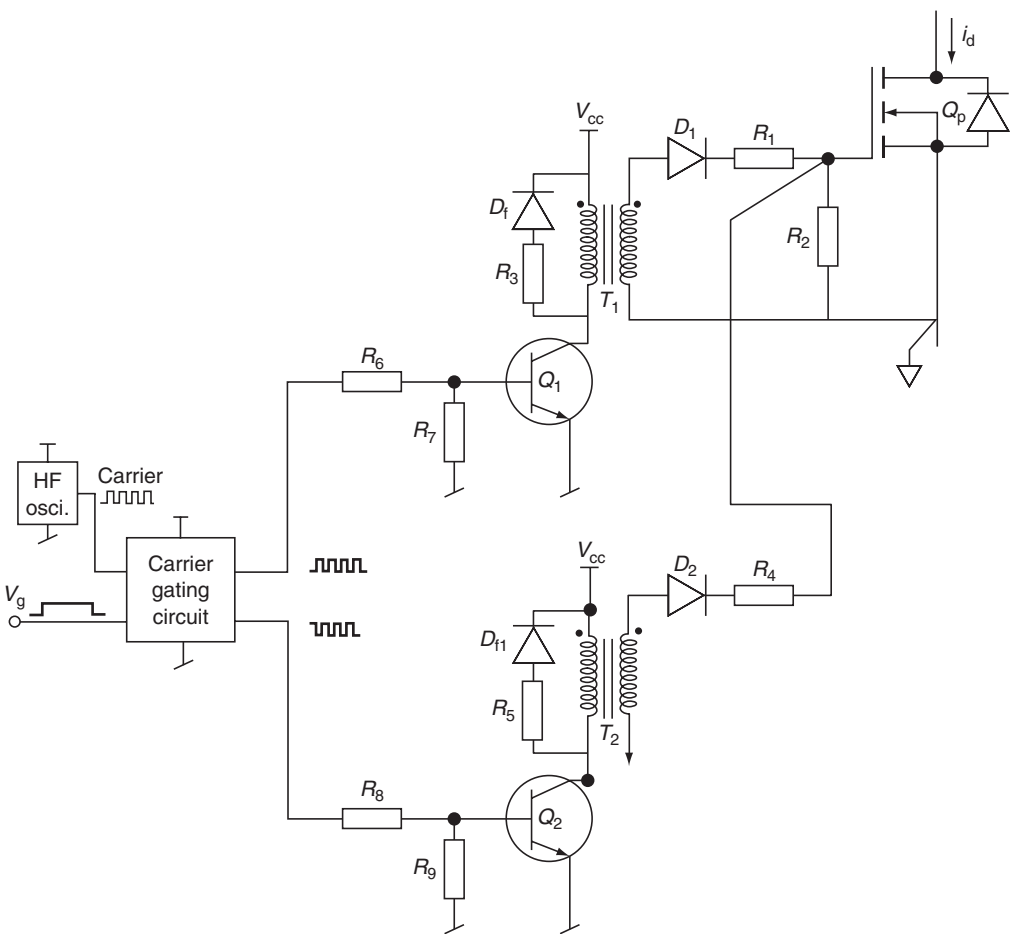


Figure 2.28 | MOSFET drive circuit-9.

switched with 180° phase difference. The secondaries of the transformers are diode ORed to obtain the desired gate pulse.

The gate drive pulse V_g is gated with a high-frequency carrier as indicated in Figure 2.28. Two switching patterns are generated to switch the two pulse transformers. One signal is obtained by directly gating the gate drive pulse V_g with the high-frequency carrier. The other signal is obtained by gating the gate drive pulse V_g with the inverted high-frequency carrier. These two signals are used to switch the two pulse transformers whose secondaries are diode ORed. In this manner, whatever be the duty cycle of the gate drive waveform V_g, the transformers are always switching at 50% duty cycle thereby avoiding transformer core saturation. In this manner, the duty cycle range of the MOSFET can be extended to 100%. One should note that, in this case, the series resistor for the MOSFET is now connected in series with each transformer before the ORing node. This is used to avoid any large circulating currents that may occur during turn-OFF of one pulse transformer and turn-ON of the other pulse transformer.

MOSFET Drive Circuit-10

A relatively simple transformer-isolated gate drive circuit with large duty cycle range capability (of 0 to almost 100%) is shown in Figure 2.29. Here, when Q_2 is switched ON, the dot poles are positive with respect to the other poles. This will charge the input capacitance of Q_p through the body diode of the MOSFET Q_1. When Q_2 is switched OFF, the dot poles become negative with respect to the other poles and will switch ON Q_1. The input capacitance of the MOSFET Q_p will discharge through Q_1 and thereby turn-OFF Q_p.

One should note that as MOSFETs do not need any gate current to maintain it in the ON-state, the pulse transformer should be designed to transfer energy just sufficient to turn-ON Q_p when Q_p is required to be turned ON and energy just sufficient to turn-ON Q_1 when Q_p is required to be turned OFF. As a consequence, the pulse transformer size will be small. The resistor R_2 is used to limit the primary current of the transformer when the transformer goes into saturation. However, one should note that in this circuit,

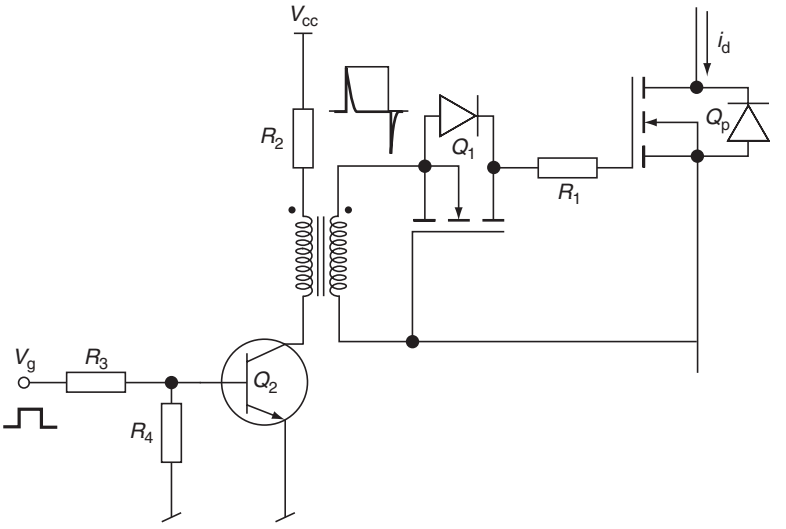


Figure 2.29 | MOSFET drive circuit-10.

when Q_p is OFF, the gate of Q_p is floating. This will leave the MOSFET open to Miller turn-ON. Any large dv/dt spikes on the drain side can easily charge up the gate–source capacitance through the drain–gate capacitance and turn-ON Q_p at an undesirable time.

MOSFET Drive Circuit-11

This gate drive circuit, which is shown in Figure 2.30, provides inherent protection against overcurrents through the drain of the MOSFET. The protection circuit discussed here can be used with any of the gate drive circuits discussed till now. Here, a transistor Q_2 is used as a series-pass element which is normally ON. When Q_2 is ON, the gate drive circuit will operate in the normal manner. If a large drain current flows through the MOSFET when it is ON, then the series-pass transistor Q_2 is immediately cut off, thereby turning OFF the MOSFET.

When Q_p is ON, the voltage across it is the ON-state drop which is $i_d R_{DS(ON)}$. When the drain current increases beyond a certain limit, the ON-state voltage will increase such that the "+" input of the comparator A_1 will be greater than the "−" input of the comparator. This will cause the output of A_1 to go high. This will enable the AND gate G_3 which will switch-OFF the PNP series-pass transistor Q_2 (output of G_2 is high at this time). The threshold limit for the ON-state voltage can be set by adjusting the reference voltage to the "−" input pin of the comparator. During the turn-ON process, when Q_p is to be turned ON, the voltage across Q_p will be high. This will cause the comparator output to be high which will, in turn, cause Q_2 to be OFF. Therefore, any gate drive pulse will not be passed on to the gate of Q_p to turn it ON. To circumvent this starting problem, a small delay of 1–5 μs is provided at turn-ON during which time the gate G_3 is forcibly

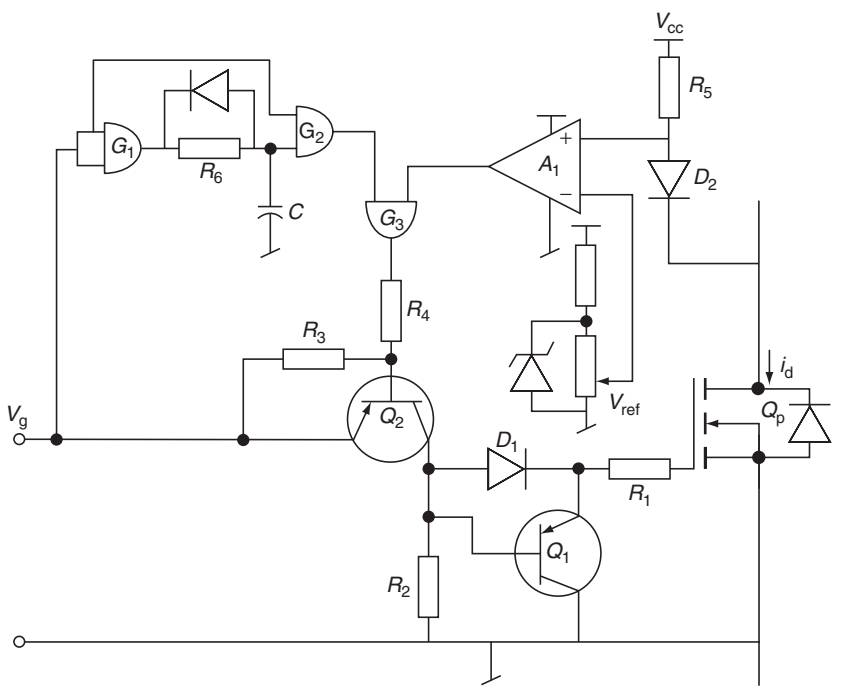


Figure 2.30 | MOSFET drive circuit-11.

disabled by the delay circuit of G_1 and G_2 such that Q_2 is ON till the MOSFET Q_p turns ON. After the initial period of 1–5 μs, G_2 will go high and enable G_3. But now A_1 output will be low as Q_p is now fully ON. A_1 will now go high only when overcurrents occur which turns OFF Q_2. This in turn will turn OFF Q_p. The gates G_1, G_2 and G_3 should be Schmitt gates in order to avoid meta-stability problems.

2.3 Snubber Circuits

In most power semiconductor devices, there is a danger of exceeding the voltage and the current ratings of the devices during the turn-OFF and turn-ON instants, respectively. As was discussed earlier in Chapter 1, when the BJT is being turned OFF, the voltage across the device V_{ce} is determined by the external circuitry. If there is an inductive load in the collector or if there is significant amount of lead inductance associated with the collector or emitter leads, then when the BJT is being turned OFF, the current through the device will fall rapidly to zero in a time corresponding to the fall time of the device. As a consequence, a large voltage spike due to $L\mathrm{d}i/\mathrm{d}t$ will occur across the device and cause the V_{ce} of the BJT to have a large spike during the fall time. This may damage the device. Therefore, it becomes essential to limit the voltage spike across the device during turn-OFF such that the voltage is within the V_{ceo} rating of the device. Similar argument applies to MOSFET and IGBT switches too. In general, during turn-OFF, the power switch should be protected against overvoltage stress. The turn-OFF voltage stresses are reduced by using circuits called the turn-OFF snubber circuits.

On the other hand, during turn-ON of the device, due to the presence of any capacitive load or parasitic capacitance across the switch, there will be a huge surge current through the device which could damage the device. Therefore, it is essential to limit the current spike through the device during turn-ON such that the device current is within the peak rating of the device. The turn-ON current stresses are reduced by using circuits called the turn-ON snubber circuits.

The snubber circuits, in general, modify the device switching characteristics and in doing so, reduce the device transient stress. In fact, the transient voltage and current stress during turn-OFF and turn-ON, respectively, are transferred to the snubber circuits. The snubber circuit action involves temporary energy storage in either an inductor or a capacitor. In re-setting these passive components, it is usual to dissipate the stored energy in a resistor as heat. As a consequence, the circuits with snubbers will be less efficient. At high frequencies, these losses may become a limiting factor because of the difficulties associated with equipment cooling. Instead of dissipating the snubber energy stored in the inductance and capacitance, alternately one may recover the energy either back into the supply or into the load. There are both passive and active circuits, available in the literature, which perform this energy recovery from the snubber. However, here the basic concepts of the snubber action will be illustrated with the generic turn-OFF and turn-ON dissipative snubber circuits. The discussion of the snubber circuits which will follow can be equally applied to BJTs, MOSFETs and IGBTs.

Turn-OFF Snubber or Shunt Snubber

Consider a power semiconductor switch Q which could be a BJT or a MOSFET or an IGBT as shown in Figure 2.31(a). Let the load be an inductive load. When the device Q is turned OFF, the load current i_L will fall to zero in a period of time corresponding to the fall time t_f. This will cause the voltage across the transistor V_{ce} to shoot up to a large value due to the $L\mathrm{d}i/\mathrm{d}t$ phenomenon where L could be the lead inductances associated with the collector and emitter leads. Therefore, it is required to limit the voltage across the collector–emitter of the transistor and it should be made to rise gradually during the period t_f as indicated in Figure 2.31(c). To slow down the change in the voltage across the device, one can connect a capacitor C

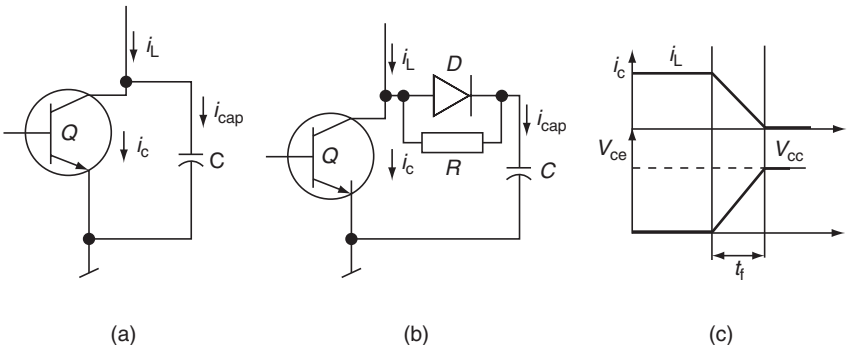


Figure 2.31 | Turn-OFF or shunt snubber.

across the device as indicated in Figure 2.31(a). As the capacitor C is connected in shunt with the device, this type of snubber is also called the shunt snubber.

Without loss of generality, one can assume that the load current is falling linearly as shown in Figure 2.31(c) during the fall time and the collector–emitter voltage across the device is rising linearly as shown in Figure 2.31(c) during the fall time when the shunt snubber is used. Referring to Figure 2.31(a),

$$i_L = i_c + i_{cap} \tag{2.22}$$

where i_c is the current through the device and i_{cap} is the current through the capacitor. During the fall time period of t_f, the current i_c through the device is given by

$$i_c = i_L\left(1 - \frac{t}{t_f}\right) \tag{2.23}$$

From Eqs. (2.22) and (2.23), the current through the capacitance is given by

$$i_{cap} = i_L - i_L\left(1 - \frac{t}{t_f}\right) = i_L\left(\frac{t}{t_f}\right) \tag{2.24}$$

The voltage v_{ce} across the device is the same as the voltage across the capacitor C. Therefore,

$$v_{ce} = \frac{1}{C}i_{cap}\mathrm{d}t \tag{2.25}$$

Use Eq. (2.24) in Eq. (2.25) and integrate within the fall time period. Then, apply the boundary condition that at the end of the fall time period, the voltage across the device or the capacitor should be V_{cc}, that is, the supply voltage. The capacitor value C is then given by

$$C = \frac{i_L t_f}{2V_{cc}} \tag{2.26}$$

One should note that though the circuit in Figure 2.31(a) will solve the problem of voltage spike during turn-OFF process of the device, it will create a serious problem during the turn-ON of the device Q. When the device Q is turned ON again, the capacitor will discharge through Q and will result in a large current

spike through the device which can damage the device. Therefore, to limit current through the device, a resistor R is introduced in series with C. The resistor R should provide the function of current limiting only during turn-ON of the device. However, during turn-OFF of the device, R is not needed. Therefore, to reduce the dissipation in R during turn-OFF, a diode is placed across R as shown in Figure 2.31(b) so that R comes into effect only during turn-ON when C discharges through R and the device.

Selection of R

When Q is turned ON, it should carry the following currents: The capacitive current discharge from C which is equal to V_{cc}/R and the load current i_L. Therefore,

$$\frac{V_{cc}}{R} + i_L < I_{cm} \tag{2.27}$$

where I_{cm} is the maximum collector-current rating of the transistor.

Re-arranging the inequality in Eq. (2.27), the following inequality is obtained:

$$R > \frac{V_{cc}}{I_{cm} - i_L} \tag{2.28}$$

It is also important to ensure that the capacitor discharges fully before the next charging when the transistor turns OFF. Therefore, there is a minimum duration of time during which time the transistor should remain ON so that the capacitor can fully discharge. The discharge time constant of the capacitor is RC. In five times this time constant, the capacitor will be almost fully discharged. Therefore

$$T_{\text{on-min}} > 5RC \tag{2.29}$$

where $T_{\text{on-min}}$ is the minimum time for which the device Q should remain in the ON-state.

Re-arranging inequality (2.29), the following inequality is obtained:

$$R < \frac{T_{\text{on-min}}}{5C} \tag{2.30}$$

From inequalities in Eqs. (2.28) and (2.30), the range for choice of R is given by

$$\frac{V_{cc}}{I_{cm} - i_L} < R < \frac{T_{\text{on-min}}}{5C} \tag{2.31}$$

One should note that when C is being charged, an energy of $CV_{cc}^2/2$ is dissipated in R and when C is being discharged, an energy of another $CV_{cc}^2/2$ is dissipated in R. Therefore, in all, an energy of CV_{cc}^2 is dissipated in R. As the function of R is to limit the current through the device Q during turn-ON only it is bypassed using a diode D as shown in Figure 2.31(b). In this case, power is dissipated in R only during capacitor discharge time. The power dissipated in R is now given by

$$P_R = \frac{1}{2} C V_{cc}^2 f_s \tag{2.32}$$

where f_s is the switching frequency of the device Q.

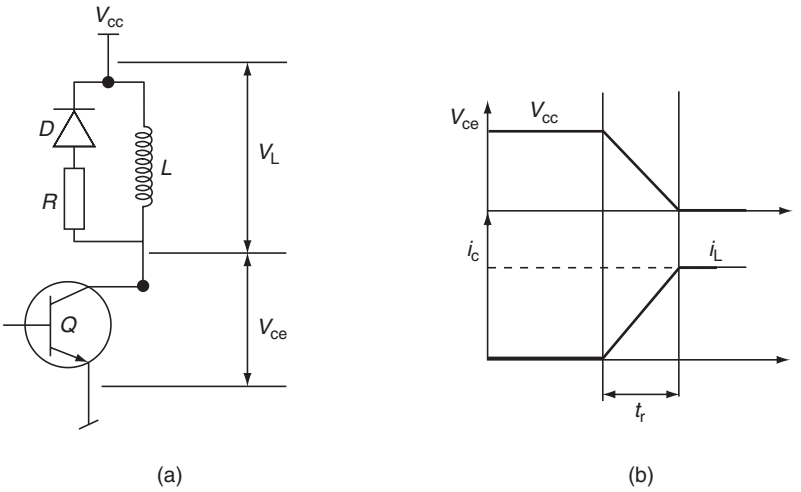


Figure 2.32 | Turn-ON or series snubber.

Turn-ON Snubber or Series Snubber

The turn-ON snubber is the dual of the turn-OFF snubber. Consider a power semiconductor switch Q which could be a BJT or a MOSFET or an IGBT as shown in Figure 2.32(a). Let the load be a capacitive load. When the device Q is turned ON, the voltage across the device (V_{ce}) will fall to zero in a period of time corresponding to the rise time t_r. This will cause the current through the transistor i_c to shoot up to a large value due to the $C(\mathrm{d}V_{cc}/\mathrm{d}t)$ phenomenon (capacitive discharge) where C could be the load capacitance or any parasitic capacitance between the collector and emitter leads. Therefore, it is required to limit the current through the transistor and it should be made to rise gradually during the period t_r as indicated in Figure 2.32(b). To slow down the change in the current through the device, one can connect an inductor L in series with the device as indicated in Figure 2.32(a). As the inductor L is connected in series with the device, this type of snubber is also called the series snubber. This snubber is an exact dual of the shunt snubber. Referring to Figure 2.32(a), diode D and resistor R provide the freewheeling path for the inductance when the transistor Q is turned OFF, so that voltage spike during turn-OFF process is limited.

Without loss of generality, one can assume that the voltage across the collector–emitter of the device Q is falling linearly as shown in Figure 2.32(b) during the rise time and the current through the device is rising linearly as shown in Figure 2.32(b) during the rise time when the series snubber circuit is used. Referring to Figure 2.32(a),

$$V_{cc} = V_{ce} + V_L \tag{2.33}$$

where V_{ce} is the voltage across the device and V_L is the voltage across the inductor. During the rise time period of t_r, the voltage V_{ce} across the device is given by

$$V_{ce} = V_{cc}\left(1 - \frac{t}{t_r}\right) \tag{2.34}$$

From Eqs. (2.33) and (2.34), the voltage across the inductor is given by

$$V_{\mathrm{L}} = V_{\mathrm{cc}}\left(\frac{t}{t_r}\right) \tag{2.35}$$

The current i_{c} through the device is the same as the current through the inductor L. Therefore

$$i_{\mathrm{c}} = i_{\mathrm{L}} = \frac{1}{L}V_{\mathrm{L}}\,\mathrm{d}t \tag{2.36}$$

Use Eq. (2.35) in Eq. (2.36) and integrate within the rise time period. Then apply the boundary condition that at the end of the rise time period, the current through the device, that is the current through the inductor, should be i_{L} (the load current value). The inductor value L is then given by

$$L = \frac{V_{cc}t_{\mathrm{r}}}{2i_{\mathrm{L}}} \tag{2.37}$$

Selection of R

When Q is turned OFF, it should withstand the following components of the voltage:

1. The voltage across the freewheeling components, that is D and R. This is equal to $i_{\mathrm{L}}R + V_{\mathrm{D}}$, where V_{D} is the diode forward drop.
2. The supply voltage V_{cc}.

Thus the following inequality should be satisfied for the device voltage rating.

$$V_{\mathrm{cc}} + i_{\mathrm{L}}R + V_{\mathrm{D}} < V_{\mathrm{ceo}} \tag{2.38}$$

where V_{ceo} is the maximum collector–emitter voltage rating of the transistor.

Re-arranging inequality (2.38), the following inequality is obtained:

$$R < \frac{V_{\mathrm{ceo}} - V_{\mathrm{cc}} - V_{\mathrm{D}}}{i_{\mathrm{L}}} \tag{2.39}$$

It is also important to ensure that the magnetic energy in the inductor L discharges fully before the next charging when the transistor turns ON. Therefore, there is a minimum duration of time during which time the transistor should remain OFF so that the inductor energy can fully discharge. The discharge time constant of the inductor is L/R. In five times this time constant, the inductor will be almost fully discharged. Therefore,

$$T_{\text{off-min}} > 5\frac{L}{R} \tag{2.40}$$

where $T_{\text{off-min}}$ is the minimum time for which the device Q should remain in the OFF-state.

Re-arranging inequality (2.40), the following inequality is obtained:

$$R > \frac{5L}{T_{\text{off-min}}} \tag{2.41}$$

From inequalities in Eqs. (2.39) and (2.41), the range for choice of R is given by

$$\frac{5L}{T_{\text{off-min}}} < R < \frac{V_{\mathrm{ceo}} - V_{\mathrm{cc}} - V_{\mathrm{D}}}{i_{\mathrm{L}}} \tag{2.42}$$

The power dissipated in R is given by

$$P_R = \frac{1}{2} L i_L^2 f_s \tag{2.43}$$

where f_s is the switching frequency of the device Q.

CONCLUDING REMARKS

In this chapter we have discussed the requirements for driving the BJTs and the MOSFETs giving a flavor for the various types of drive circuits. It should, however, be noted that by no means is the list of discussed drive circuits exhaustive. There are many ICs and hybrid circuits available commercially that implement the drives circuits of all the switches of either half-bridge or full-bridge or three-phase full-bridge as a single device. These integrated devices in addition provide features like overcurrent protection and thermal protection. If a fault occurs, a fault output is provided that can be used as an interrupt signal to disable the drive pulses to the bridge circuit used in a converter or inverter application. Whether one uses a discrete drive circuit or an integrated drive circuit device for a specific application depends on the cost, size, isolation feature, protection features and interface compatibility to microcontrollers or DSPs. However it should be borne in mind that the reliability and the performance of the power switch is very much dependent on the drive circuit. It is not without reason that the phrase "*a power electronic product is as good as the drive circuit*" is central to the performance of power electronic systems.

BJTs are generally more robust during handling whereas care must be taken in handling MOSFETs and IGBTs. The MOSFETs and IGBTs are voltage-controlled devices. Their turn-ON depends on the gate charge that is provided. The body of a person contains sufficient static charge to charge up the gate capacitance to a few hundred volts whereas the gate–source or gate–emitter can handle only around 20 V. Therefore, in many cases, the devices will get damaged just by handling. They must be place on anti-static pads with the gate and source/collector shorted. In addition, a protective 15 V Zener diode is connected between gate and source/emitter to provide a measure of clamping if the static induced voltage exceeds 15 V. However, once the MOSFETs and IGBTs are mounted onto the printed wiring boards or connected to the drive circuits, they are usually very robust.

The key to reliable power switch performance is to ensure that the locus of the operating point is always within the forward and reverse safe operating areas (SOARs). Any transition of the operating point locus across the boundary of the SOAR will cause overstressing of the semiconductor bulk and the junctions. This will cause fast aging of the device leading to deterioration and premature failure. Most failures of the power switches are due to non-restriction of the operating point locus within the SOAR during switching transitions. If such a situation is even suspected, then appropriate snubber circuits must be incorporated to reduce the device stresses even though it may be at the expense of efficiency.

Practice and practical are the essence of a good engineer. One must strive to both simulate and breadboard the BJT and MOSFET drive circuits discussed in this chapter and more by referring to literature. The next section provides few exercises that can be simulated in spice and also implemented by hardware breadboarding. The insights gained in obtaining experimental clarifications will be helpful while designing DC–DC converters and DC–AC inverters.

LABORATORY EXERCISES

1. Consider the BJT drive test circuit shown in Figure 2.33. It consists of the test transistor Q_p that is to be studied. Q_p is used to switch a resistive load R_c that draws power from a 15 V DC supply. The base drive circuit comprises complementary transistors Q_1 (NPN) and Q_2 (PNP) transistors that are connected as shown. Q_1 collector is connected to V_{cc} positive pole and Q_2 collector is connected to V_{ee} negative pole as shown. The base pulse signal source V_b is used to provide the base drive pulse signal to switch the power transistor Q_p at a specific frequency.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Hardware breadboarding

Tasks for study:

- **(a)** Rig up the circuit/netlist as given in Figure 2.33.
- **(b)** Set V_{cc} = 10 V, V_{ee} = 5 V. Measure and tabulate i_c, i_b and V_{ce} of Q_p for various values of R_1.
- **(c)** From the tabulated values of i_b, i_c and V_{ce}, compute the product $i_c \times V_{ce}$ the instantaneous power loss in the device. Plot i_b, i_c, V_{ce} and $i_c \times V_{ce}$ versus time.
- **(d)** What is the effect of R_1 on the above waveforms?
- **(e)** At what value of i_c and i_b does V_{ce} attain saturation value of around 0.3 V?
- **(f)** At what value of i_c and i_b does V_{ce} attain a value around 0.7 V which is just at the boundary of saturation?
- **(g)** Set R_1 to an appropriate value. Set V_{ee} = 5 V. Measure and tabulate i_c, i_b and V_{ce} of Q_p for various values of V_{cc}.
- **(h)** From the tabulated values of i_b, i_c and V_{ce}, compute the product $i_c \times V_{ce}$ the instantaneous power loss in the device. Plot i_b, i_c, V_{ce} and $i_c \times V_{ce}$ versus time.
- **(i)** What is the effect of V_{cc} on the above waveforms?
- **(j)** Set R_1 to an appropriate value. Set V_{cc} = 10 V. Measure and tabulate i_c, i_b and V_{ce} of Q_p for various values of V_{ee}.

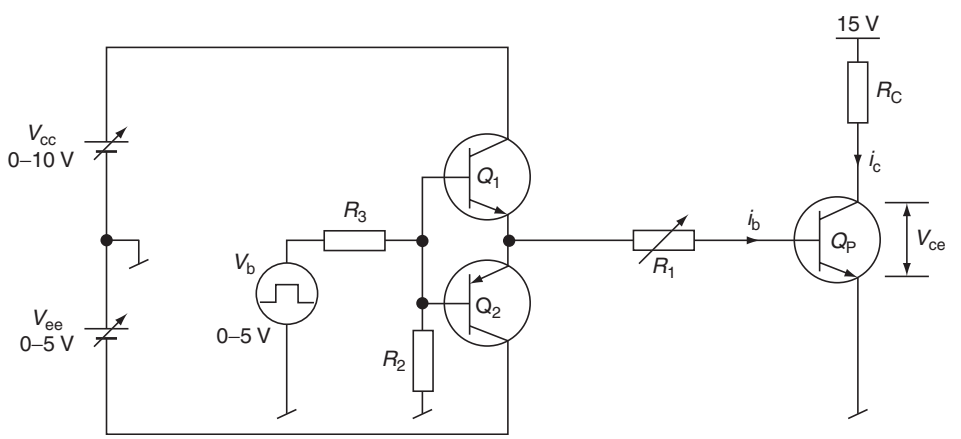


Figure 2.33 | BJT drive test circuit.

(k) From the tabulated values of i_b, i_c and V_{ce}, compute the product $i_c \times V_{ce}$ the instantaneous power loss in the device. Plot i_b, i_c, V_{ce} and $i_c \times V_{ce}$ versus time.

(l) What is the effect of V_{ee} on the above waveforms?

2. Consider the BJT drive test circuit shown in Figure 2.34. It is the same as the test circuit of Figure 2.33 but for the inclusion of the speed up circuit.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice

b. Hardware breadboarding

Tasks for study:

(a) Rig up the circuit/netlist as given in Figure 2.34.

(b) Set V_{cc} = 10 V, V_{ee} = 5 V. Set R_1 to provide steady-ON base-current value appropriate for the collector current. Measure and tabulate i_c, i_b and V_{ce} of Q_p for various values of R_2. Ensure in all cases that R_2 is less than $R_1/5$.

(c) From the tabulated values of i_b, i_c and V_{ce}, compute the product $i_c \times V_{ce}$ the instantaneous power loss in the device. Plot i_b, i_c, V_{ce} and $i_c \times V_{ce}$ versus time.

(d) What is the effect of R_2 on the above waveforms?

(e) Set V_{cc} = 10 V. Set R_1 to provide steady-ON base-current value appropriate for the collector current. Set R_2 to provide the i_{b+}. Measure and tabulate i_c, i_b and V_{ce} of Q_p for various values of V_{ee}.

(f) From the tabulated values of i_b, i_c and V_{ce}, compute the product $i_c \times V_{ce}$ the instantaneous power loss in the device. Plot i_b, i_c, V_{ce} and $i_c \times V_{ce}$ versus time.

(g) What are the values of i_{b+} and i_{b-} at which turn-ON and turn-OFF is best and switching power dissipation is least?

(h) On incorporating another BJT in Darlington configuration with the output power BJT Q_p as shown in Figure 2.35, what is the effect on the base drive requirements?

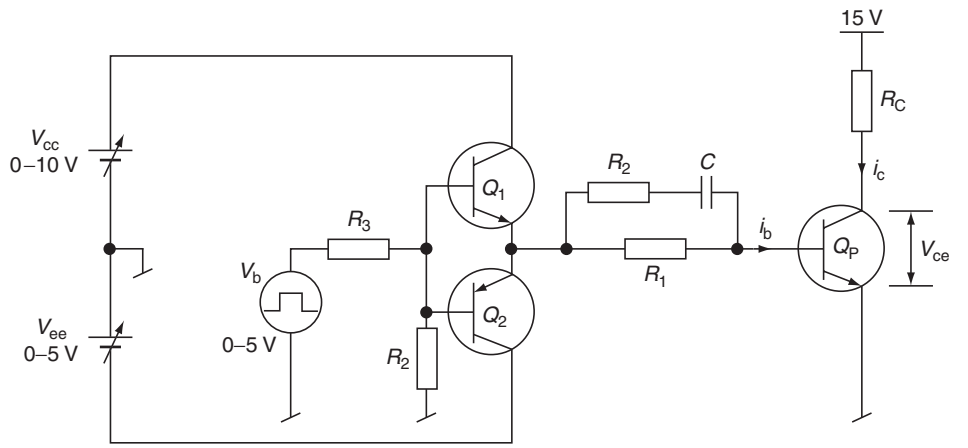


Figure 2.34 | BJT drive test circuit.

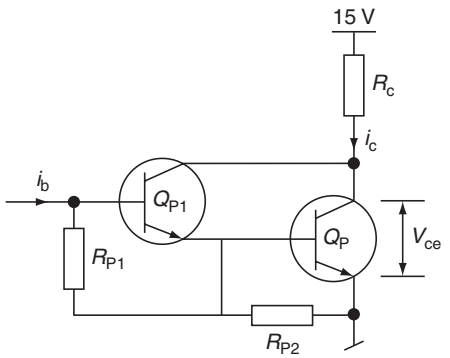


Figure 2.35 Darlington configuration of the output power transistor. (All other portions of the drive are the same as in Figure 2.34.)

3. Consider the MOSFET drive test circuit shown in Figure 2.36. It consists of the test MOSFET Q_p that is to be studied. Q_p is used to switch a resistive load R_c that draws power from a 15 V DC supply. The gate drive circuit comprises complementary transistors Q_1 (NPN) and Q_2 (PNP) transistors that are connected as shown. Q_1 collector is connected to V_{cc} positive pole and Q_2 collector is connected to V_{ee} negative pole as shown. The gate pulse signal source V_g is used to provide the gate drive pulse signal to switch the power MOSFET Q_p at a specific frequency.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Hardware breadboarding

Tasks for study:

- **(a)** Rig up the circuit/netlist as given in Figure 2.36.
- **(b)** Set V_{cc} = 15 V, V_{ee} = 5 V. Measure and tabulate i_d, i_g and V_{ds} of Q_p for various values of R_1.
- **(c)** What is the effect of R_1 on the turn-ON and turn-OFF times? From the tabulated values of i_d, i_g and V_{ds} estimate the gate charge required to turn-ON the MOSFET to full enhancement.
- **(d)** Set R_1 to an appropriate value. Set V_{cc} = 15 V. Measure and tabulate i_d, i_g and V_{ds} of Q_p for various values of V_{ee}.
- **(e)** What is the effect of V_{ee} on the above waveforms?

4. Consider the MOSFET shunt snubber test circuit shown in Figure 2.37. It consists of the test MOSFET Q_p that is used to switch a resistive load R_d that draws power from a DC supply, V_{dd}. The inductance L_1 indicates the track inductance of the inductive effects in the load. The shunt snubber circuit comprising R, C and D is connected

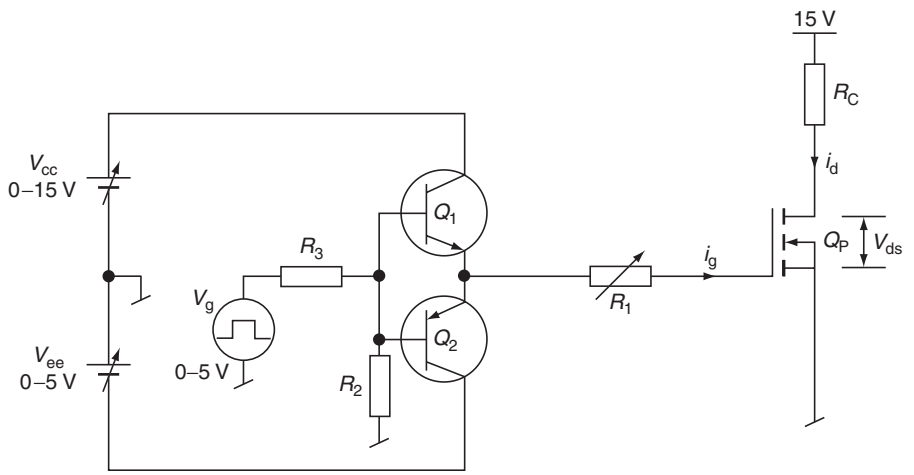


Figure 2.36 MOSFET drive test circuit.

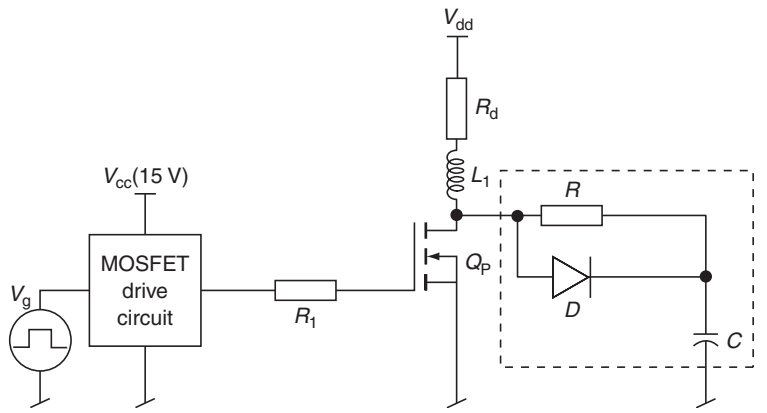


Figure 2.37 | MOSFET shunt snubber test circuit.

as shown. The gate pulse signal V_g is used to provide the gate pulse signal to switch the power MOSFET Q_p at a specific frequency. The pulse signal from V_g is passed through an appropriate MOSFET drive circuit to provide the necessary gate charge to switch the MOSFET Q_p.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Hardware breadboarding

Tasks for study:

(a) Rig up the circuit/netlist as given in Figure 2.37.

(b) Set V_{dd} = 100 V, R_d = 20 Ω. Measure and tabulate i_d and V_{ds} of Q_p for various values of C.

(c) What is the effect of C on the V_{ds} waveform?

(d) What is the effect of R on the turn-ON drain current of the MOSFET?

(e) Measure the rms current through R and compute the power dissipation in R with and without D.

(f) How does the modified shunt snubber circuit as shown in Figure 2.38 operate? What is the effect on the i_d and V_{ds} waveforms? What is the role of RC time constant on the V_{ds} waveform? What should be the value of R and C for a given load?

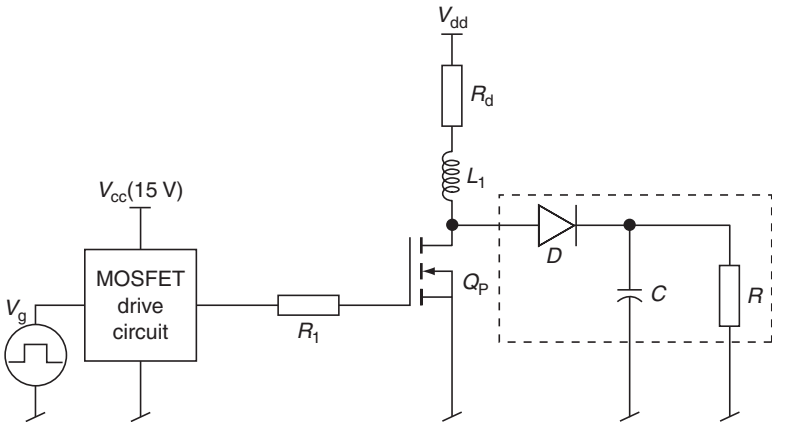


Figure 2.38 | Modified shunt snubber circuit.

5. Consider the MOSFET series snubber test circuit shown in Figure 2.39. It consists of the test MOSFET Q_p that is used to switch a capacitive load. The series snubber circuit comprising R, L and D is connected as shown. The gate pulse signal V_g is used to provide the gate pulse signal to switch the power MOSFET Q_p at a specific frequency. The pulse signal from V_g is passed through an appropriate MOSFET drive circuit to provide the necessary gate charge to switch the MOSFET Q_p.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware breadboarding

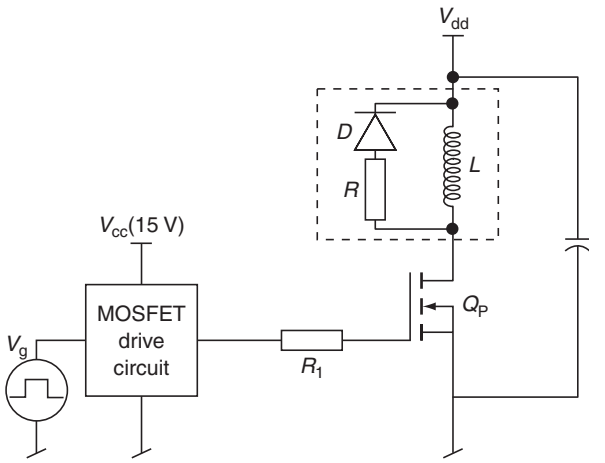


Figure 2.39 | MOSFET series snubber test circuit.

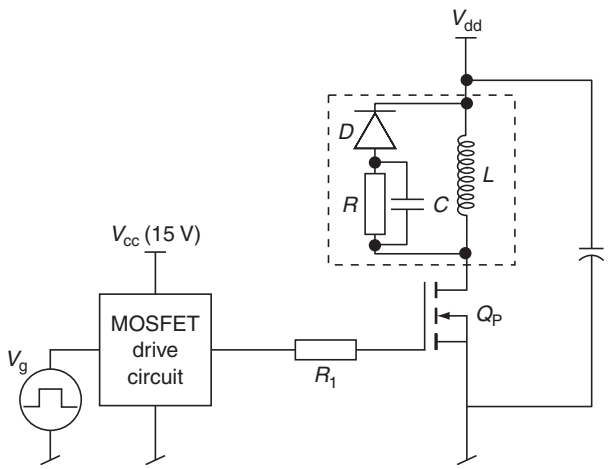


Figure 2.40 | Modified series snubber circuit.

Tasks for study:

(a) Rig up the circuit/netlist as given in Figure 2.39.

(b) Set V_{dd} = 15 V. Measure and tabulate i_d and V_{ds} of Q_p for various values of L.

(c) What is the effect of L on the i_d waveform?

(d) What is the effect of R on the turn-OFF drain–source voltage of the MOSFET?

(e) Measure the rms current through R and compute the power dissipation in R.

(f) How does the modified series snubber circuit as shown in Figure 2.40 operate? What is the effect on the i_d waveform and V_{ds} waveform? What is the role of RC time constant on the i_d and V_{ds} waveforms? What should be the value of R and C for a given load?

FILL IN THE BLANKS

1. BJT is a _____ device.

2. In BJT switching, the base currents are _____ on the collector currents.

3. During the turn-ON period, the base current must be _____ than the final or steady-ON value.

4. During the turn-OFF period, a _____ base current should be provided to quickly remove the stored charges in the transistor.

5. For the BJT drive circuit-1, R_2 is used to provide the _____ path for removing the _____ charges during turn-OFF.

6. In BJT drive circuit-2, the two transistor stages, Q_2 and Q_1, provide the _____ gain to suitably drive Q_p.

7. In BJT drive circuit-3, the turn-OFF time of the power transistor is improved by using an _____.

8. During turn-OFF, the negative base current for fast turn-OFF in BJT drive circuit-6 is given by _____.

9. If the transistor is operated to be just out of _____, then the stored base charge is small and the turn-OFF times will be very fast.

10. If the collector current _____, then the transistor will be overdriven into saturation for the same base drive.

11. If the collector current _____, then the transistor will come out of saturation and operate in the active region.

12. In applications where the load current can vary in a wide range, the base current should vary _____ to the collector current. Such a base drive circuit is called _____.

13. Isolation between the higher power collector side and the lower power base side of the power transistor is provided either by using _____ or by using _____.

14. In the case of opto-coupler isolation, the duty cycle can be varied from 0 to 1, but _____ power supplies for the diode side and transistor side of the opto-coupler are required.

15. In the case of transformer isolation, the duty cycle can be varied only from _____.

16. MOSFET is a _____.

17. In MOSFET switching, the gate currents are _____ on the drain currents.

18. The gate current is _____ during the steady-ON period when the MOSFET is ON.

19. Compared to the BJT, the MOSFET drive circuit power requirement is _____.

20. The benefits of the dual transformer ORed circuit as given in MOSFET drive circuit-9 are _____, _____ and _____.

21. In MOSFET drive circuit-10, when the output power MOSFET is OFF, the gate is _____ and open to _____ due to large dv/dt spikes on the drain side.

22. If there is an inductive load in the collector or if there is significant amount of lead inductance associated with the collector or emitter leads, then when the BJT is being turned OFF, the current through the device will fall rapidly to zero in a time corresponding to the fall time of the device leading to large _____ stress on the device.

23. _____ limits the voltage across the device during turn-OFF process.

24. Turn-OFF snubber is also called _____.

25. If there is a capacitive load or parasitic capacitance across the device, then a large _____ flows through the device at turn-ON.

26. Turn-ON snubber circuit limits the _____ the device during turn-ON process.

27. Turn-ON snubber is also called _____.

28. The circuits with snubbers will be _____ efficient.

DESCRIPTIVE QUESTIONS

1. Discuss the effect of the base drive waveshape on the turn-ON and turn-OFF speeds of the device.

2. Discuss the effect of the base drive waveshape on the switching power dissipation.

3. What are the requirements of a good base drive? Illustrate with the base-current waveform.

4. Explain the operation of the BJT base drive circuit-1.

5. Explain the operation of the BJT base drive circuit-2.

6. How does the operation of BJT drive circuit-3 differ from that of BJT drive circuit-4?

7. In BJT drive circuit-5, the R_2-C is the speed up circuit that ensures fast turn-ON. Explain?

8. Explain the difference between the BJT drive circuit-5 and BJT drive circuit-6.

9. What is Baker's clamp? Where and why is it used?

10. Explain the operation of a non-isolated proportional-base drive circuit.

11. What is the function of D_1 in the proportional-base drive circuit as given in BJT drive circuit-7?

12. Explain the operation of the opto-coupler-based BJT drive circuit-8.

13. Explain the operation of the transformer-based BJT drive circuit-9.

14. Explain the operation of the BJT drive circuit-12.

15. What is the difference between BJT drive circuit-12 and BJT drive circuit-13?

16. Explain the operation of MOSFET drive circuit-1. What is the function of resistor R_2? Under what constraints should R_2 be used?

17. In MOSFET drive circuit-2, why are the buffers connected in parallel?

18. Explain the operation of MOSFET drive circuit-3.

19. Explain the difference between the opto-isolated BJT drive circuit-8 and MOSFET drive circuit-4.

20. "*The MOSFET drive circuit-6 is used to drive the top MOSFET of bridge arms.*" Explain its operation.

21. How can the duty cycle range be improved with transformer isolation between the collector side and the base drive side?

22. In the MOSFET drive circuit-9, what is the function of the secondary-side resistors R_1 and R_4?

23. Explain the operation of MOSFET drive circuit-10.

24. The protection circuit discussed in MOSFET drive circuit-11 needs to be disabled at turn-ON. Why?

PROBLEMS

1. A BJT has to switch a load of 1 A. The base drive power is derived from 5 V voltage source. Calculate the values of resistors R_1 and R_2 for the BJT having saturation h_{FE} as 100.

2. A BJT has to switch a load of 10 A that is connected to a 400 V DC link/bus. The BJT drive circuit-3 is used to drive the power transistor. The power transistor has a saturation h_{FE} of 80. The base drive is powered from a 10 V DC source. Calculate the values of R_1, R_2 and L.

3. A load of 10 A is to be switched by a BJT. The BJT base drive circuit is powered from 10 V DC source. The BJT selected for the application has the following specifications: $h_{FE(sat)}$ = 100, t_r = 2 μs and t_f = 2 μs. If the BJT drive circuit-5 is used in the application, then calculate the values of R_1, R_2 and C.

4. In BJT drive circuit-10, consider the time period when the transistor Q_1 is ON. What is the voltage across the freewheeling diode D_f during this time period?

5. In the BJT drive circuit-11, the PNP transistor Q_1 is used to ensure fast turn-OFF of the power transistor. What should be the V_{ce} and the collector-current rating of the transistor Q_1?

6. For a varying load current that has a maximum value of 10 A, design a proportional-base drive circuit as discussed in BJT drive circuit-13.

7. A particular MOSFET requires 400 nC to take it to full enhancement. The MOSFET is driven from a 15 V source. If the peak gate current during turn-ON is 2 A, then calculate the turn-ON time.

8. For Problem 7 above, what is the peak current requirement for a turn-ON time of 1 μs? Calculate the gate resistance that needs to be connected in series.

9. What should be the voltage and current rating of Q_2 in MOSFET drive circuit-3?

10. A bridge arm is used as part of a converter. The DC-link voltage is 400 V. The gate power supply is derived from a 15 V power supply. The MOSFET drive circuit-6 is employed to drive the top MOSFET of the bridge arm. The bridge MOSFETs are switching a drain current of 10 A. The $R_{DS(ON)}$ of the bridge MOSFETs is 0.1 Ω. When the top MOSFET is ON, what is the reverse voltage across the diode D?

11. In the protection circuit of MOSFET drive circuit-11, a drain current of 20 A flows through the power MOSFET that has a $R_{DS(ON)}$ of 0.1 Ω. What is the voltage at the "+" terminal of amplifier A_1?

12. A MOSFET is switching a 20 A inductive load from a 100 V DC source. The fall time of the device is 0.5 μs. Calculate the shunt snubber capacitance value.

13. For Problem 12, the peak current rating of the MOSFET is specified as 30 A. The converter, where the MOSFET is used, operates at a switching frequency of 20 kHz. The range of duty cycle is from 0.2 to 0.7. What is the value

of the shunt snubber resistor? What is the power dissipation in the shunt snubber resistor?

14. For Problem 13, what is the snubber diode current and peak inverse voltage rating?

15. An IGBT is switching a capacitive load. The load current is 20 A drawn from a 100 V DC source. To protect against the turn-ON current surges, a series snubber is used. The rise time of the device is 2 μs. Calculate the series snubber inductance value. The voltage rating of the device is specified as 150 V. The converter where the IGBT is used operates at a switching frequency of 20 kHz. The range of duty cycle is from 0.2 to 0.7. What is the value of the series snubber resistor? What is the power dissipation in the shunt snubber resistor?

ANSWERS

Fill in the Blanks

1. current-controlled
2. dependent
3. higher
4. negative
5. discharge; stored base
6. current
7. inductor
8. $(V_{cc} + V_{besat})/R_1$
9. saturation
10. decreases
11. increases
12. in proportion; proportional-base drive circuit
13. opto-couplers; transformers
14. isolated
15. 0 to 0.5
16. voltage-controlled device
17. not dependent
18. zero
19. lesser
20. isolation; full range of duty cycle; no requirement for isolated power supplies on the secondary side
21. floating; turn-ON
22. voltage
23. Turn-OFF snubber circuit
24. shunt snubber
25. current surge
26. current through
27. series snubber
28. less

CHAPTER 3

Rectifiers

Learning Objectives

After reading this chapter, you will be able to:

- understand the operation of rectifiers and the effect of the various loads on rectifier functioning.
- draw current and voltage waveforms at various points in the circuit.
- understand and design capacitor-filter rectifier circuits.
- understand the operation of controlled rectification.

This chapter deals primarily with converting the AC voltage that is available from the mains to the DC voltage which is required for most electronic products. The AC voltage is first converted to a pulsating DC voltage by using diodes or thyristor. This pulsating DC is then filtered to provide smooth DC voltage. The rectifiers may be broadly classified as

1. uncontrolled rectifiers;
2. controlled rectifiers.

The uncontrolled rectifiers use diode as the semiconductor power switch. As the turn-ON and turn-OFF of the diode is uncontrolled (discussed in Chapter 1), such rectifier topologies using diodes as the power switch are called uncontrolled rectifier. On the other hand, if any of the controllable power semiconductor switches like the bipolar junction transistors (BJTs), metal oxide semiconductor field effect transistors (MOSFETs), insulated gate bipolar transistors (IGBTs) and thyristors are used, then such rectifiers are called controlled rectifiers.

The uncontrolled rectifiers are discussed first in this chapter. The various single- and three-phase rectifier configurations are discussed. This is followed by a discussion on the popular capacitor input filter rectifier wherein the operation and design of capacitor input filters are dealt. The inrush current at rectifier startup is also addressed by including some methods to solve this problem. The discussion on the capacitor input filter is followed by the inductance input filter or the LC-filter rectifier. The output ripple content of the LC filter is analyzed. The effect of LC filter on the currents in the rectifier is also discussed.

After discussing uncontrolled rectifiers, the controlled rectifiers are addressed. The phase-controlled converters, viz., single-phase half-wave controlled converters, single-phase full-wave converters, three-phase half-wave controlled converters, three-phase full-wave half-controlled converters and three-phase full-wave full-controlled converters are considered. The firing sequences and the corresponding waveforms at the rectifier outputs are also discussed.

3.1 Uncontrolled Rectifiers

Electrical power generation and distribution is usually accomplished in the form of AC currents and voltages. However, many types of electronic and electrical equipments operate from DC sources. The AC voltage must therefore be rectified and filtered to provide a DC output voltage at a specified current or power level. Depending on the load power requirements, the AC input may be obtained from

1. single-phase supply for low-to-medium power requirements;
2. three-phase supply for medium-to-high power requirements.

A transformer may be used in between the AC input supply and the rectifier input to provide

1. voltage scaling;
2. isolate the input from the rectifier output.

In such cases, the transformer should be capable of handling the entire load power and should be designed for the mains frequency. This would increase the cost, size and weight of the rectifier. Frequently, the input voltage is directly rectified by using a transformer and the filtered rectifier output is then switched at a high frequency across a load or transformer supplying the load. A typical example of the latter case is switching power supplies and step-up frequency converters. In either case, the magnetics operate at a higher frequency for size, weight and cost reduction. These are classified as DC–DC converters that are discussed in Chapter 5.

In high-power systems, regulation is frequently accomplished by controlling the phase or conduction time of the AC wave by thyristors (silicon-controlled rectifiers; SCRs). In low-power systems, such as AC–DC power supplies, regulation is usually achieved by either a transistor post regulator like in linear regulator or by pulse-width modulation for switched-mode regulation.

3.2 Rectifier Circuits

The uncontrolled rectifier circuits consist primarily of diode semiconductor switches. The diodes are connected in various topological configurations. The source for the rectifier circuits is considered to be a sinusoidal voltage source such as the mains/grid that is the most common source in many applications. Based on the type of the input source for the rectifiers, they are broadly classified as

1. single-phase rectifier circuits;
2. three-phase rectifier circuits.

As the names imply, the single-phase rectifier circuit topologies are designed for applications wherein the input source is the single-phase grid and likewise, the three-phase rectifier circuit topologies are designed for applications with three-phase grid as the input source.

Single-Phase Circuits

In this category of rectifier topologies, there are three basic rectifier configurations that are popular:

1. half-wave rectifier;
2. full-wave center-tapped rectifier;
3. full-wave bridge rectifier.

These three basic configurations will now be discussed.

Half-Wave Rectifier

The single-phase half-wave rectifier circuits for various load conditions are shown in Figure 3.1. The rectifier topology of Figure 3.1(a) is applicable for resistive and/or capacitive loads. On the other hand, the rectifier topology of Figure 3.1(b) is applicable for inductive and/or resistive–inductive loads. The diode D_2 across the load is needed for providing a path for the trapped kinetic energy in the inductor to discharge when the diode D_1 is reverse-biased. As the diode D_2 provides a path for the inductor current to freewheel through it when D_1 is OFF, D_2 is called the *freewheeling diode*. The mains/grid voltage V_{in} is applied at the primary of the transformer. The voltage at the secondary of the transformer is nV_{in}. The choice of the turns ratio n depends on the load voltage requirements.

Whenever the dot poles of the transformer are positive, diode D_1 will conduct and the load voltage V_L will follow the secondary voltage V_s. During this time the diode D_2 in the case of circuit of Figure 3.1(b) will be reverse-biased and therefore OFF. When V_s goes negative with respect to the dot pole, D_1 will be reverse-biased and will switch OFF. If the load is inductive, then the inductive current will freewheel through diode D_2.

Figure 3.2 shows the voltage and current waveforms for a resistive load. In this case the diode D_2 is not operative. Referring to the waveforms shown in Figure 3.2, the load voltage is a pulsating half-sinusoid as indicated. It is given by

$$V_{Lav} = \frac{nV_m}{\pi} \tag{3.1}$$

where V_{Lav} is the average voltage across the load resistor R_L; V_m is the peak voltage of the input sine wave; n is the transformer turns ratio.

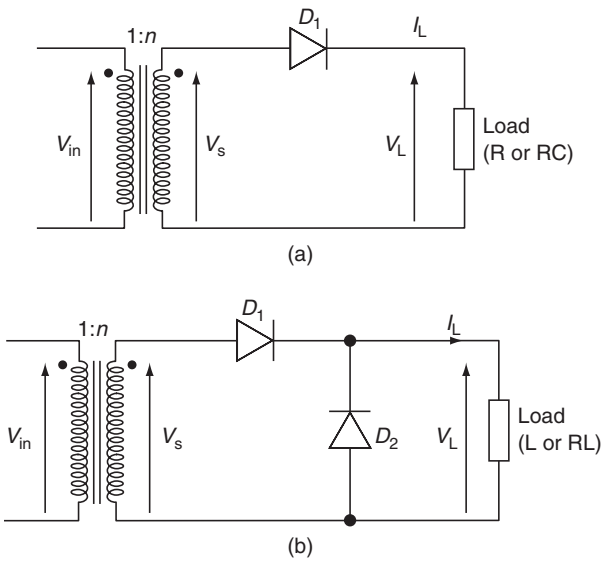


Figure 3.1 Half-wave rectifier configuration: (a) For resistive–capacitive loads; (b) for resistive–inductive loads.

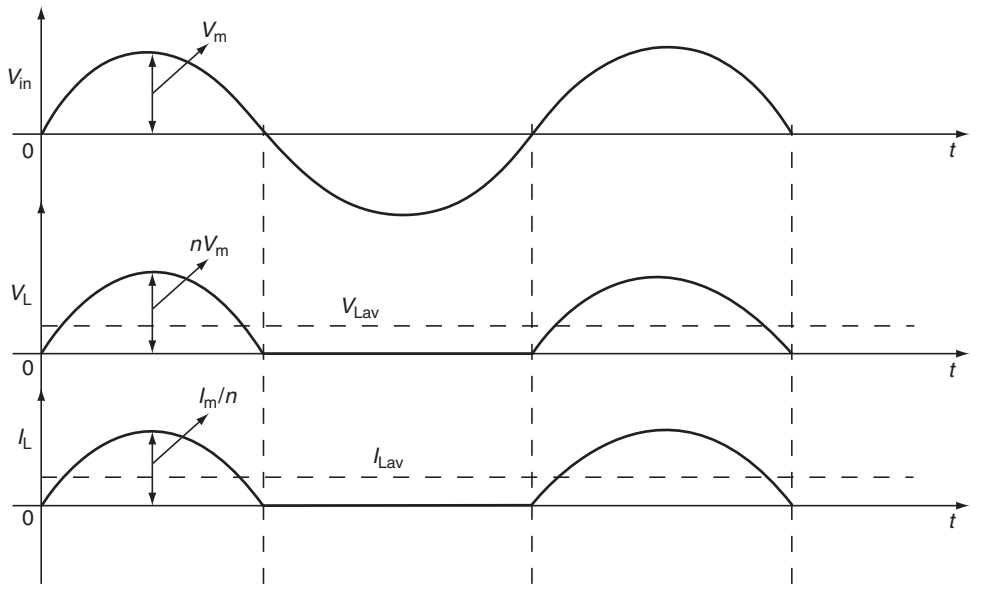


Figure 3.2 | Half-wave rectifier: Waveforms for resistive load.

The average load current is given as

$$I_{Lav} = \frac{nV_m}{\pi R_L} \tag{3.2}$$

where I_{Lav} is the average load current and the remaining symbols have their usual meaning. The power delivered to the load is given as

$$\begin{aligned} P_o &= \frac{1}{T}\int_0^{T/2} (nV_m \sin\omega t)\left(\frac{nV_m \sin\omega t}{R_L}\right) dt \\ &= \frac{n^2 V_m{}^2}{4R_L} = \left(\frac{nV_m}{2}\right)\left(\frac{nV_m}{2R_L}\right) \\ &= V_{Lrms} I_{Lrms} \end{aligned} \tag{3.3}$$

where V_{Lrms} is the root mean square (rms) voltage across the load resistor and I_{Lrms} is the rms current flowing through the load resistor.

The current through the secondary of the transformer is also same as that of the load and hence the rms secondary current is given as

$$I_{srms} = I_{Lrms} \tag{3.4}$$

The primary or line side rms current is given as

$$I_{prms} = nI_{srms} \tag{3.5}$$

It can be seen from the waveforms of Figure 3.2 that the maximum peak inverse voltage (PIV) that D_1 should withstand when OFF is V_m. Likewise in the case of inductive loads, D_2 is operative. When D_2 is OFF, D_1 is ON. The maximum PIV that D_2 should withstand when OFF is also V_m.

The output voltage is only a half sine wave. Some form of output filtering is essential to obtain a low ripple output voltage. This topology can also be used for direct line rectification by removing the costly transformer. However, the half-wave circuit is usually limited to low-power, poor output ripple applications. This topology is used to obtain the DC bus voltage for inputs of linear regulators in power supplies.

Full-Wave Center-Tapped Rectifier

The full-wave center-tapped rectifier circuit is shown in Figure 3.3. Here the use of a transformer is mandatory. The turns ratio n is defined as the ratio of the number of windings in the secondary between the dot pole and the center-tap to the number of windings in the primary. Referring to Figure 3.3, diode D_1 will conduct when the dot poles of the transformer are positive and V_L will follow the positive half of V_s. When the dot poles are negative, diode D_2 will conduct resulting in a full wave rectified waveform at the output. Here, the peak-reverse voltage for the diodes is twice V_s, thus demanding a higher reverse voltage rating for a given V_L requirement.

In the case of the center-tapped rectifier circuit, there is no need for a separate freewheeling diode. This is due to the fact that one of either D_1 or D_2 is always ON. The voltage and current waveforms are shown in Figure 3.4.

Referring to the waveforms shown in Figure 3.4, the load voltage is a pulsating full sinusoid as indicated. It is given by

$$V_{Lav} = \frac{2nV_m}{\pi} \tag{3.6}$$

where V_{Lav} is the average voltage across the load resistor R_L; V_m is the peak voltage of the input sine wave; n is the transformer turns ratio. The average load current is given as

$$I_{Lav} = \frac{2nV_m}{\pi R_L} \tag{3.7}$$

where I_{Lav} is the average load current.

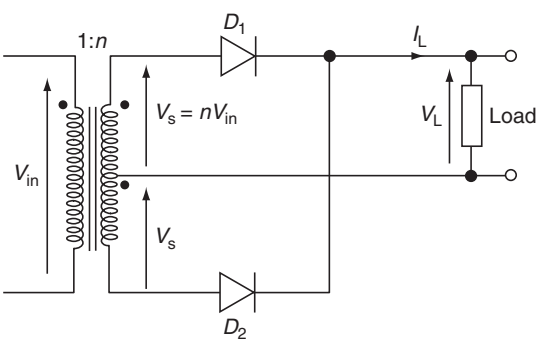


Figure 3.3 | Full-wave center-tapped rectifier.

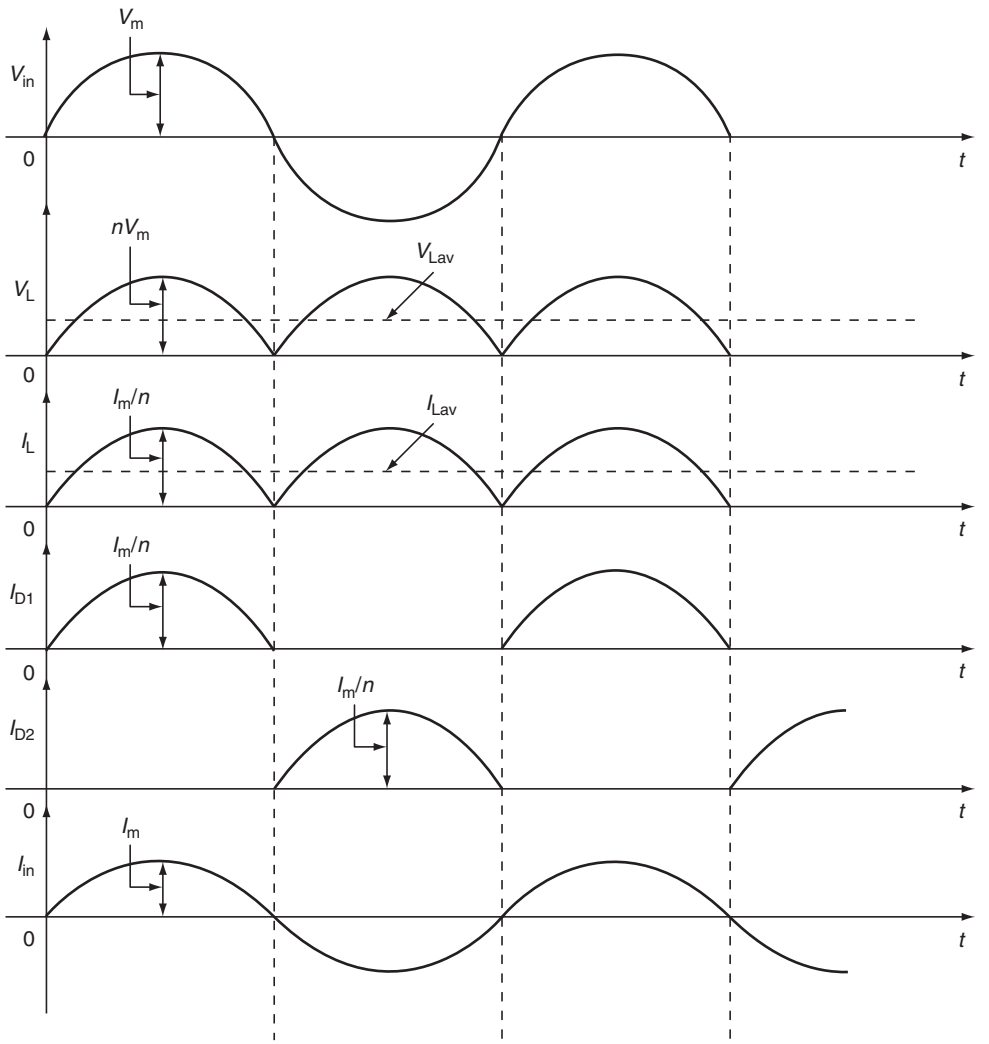


Figure 3.4 | Center-tapped full-wave rectifier: Waveforms for resistive load.

The power delivered to the load is given as

$$P_o = \frac{2}{T}\int_0^{T/2}\left(nV_m \sin\omega t\right)\left(\frac{nV_m \sin\omega t}{R_L}\right)dt$$
$$= \frac{n^2 V_m^{\ 2}}{2R_L} = \left(\frac{nV_m}{\sqrt{2}}\right)\left(\frac{nV_m}{\sqrt{2}R_L}\right) = V_{Lrms} I_{Lrms} \tag{3.8}$$

where V_{Lrms} is the rms voltage across the load resistor and I_{Lrms} is the rms current flowing through the load resistor.

The current through the secondary of the transformer is a half-wave rectified waveform. The portion of the transformer secondary winding above the center-tap carries current only when D_1 is ON as indicated in I_{D1} waveform of Figure 3.4. On the other hand, the portion of the secondary winding below the center-tap carries current only when D_2 is ON as indicated in I_{D2} waveform of Figure 3.4. Thus the secondary winding currents are half-wave rectified waveforms. The rms value of the secondary winding current is given as

$$I_{srms} = \frac{nV_m}{2R_L} = \frac{I_{Lrms}}{\sqrt{2}} \tag{3.9}$$

The primary or line side current is however a full sine waveshape as this is the algebraic sum of the reflected secondary winding currents. This is indicated in the I_{in} waveform of Figure 3.4. The rms primary or line side current is given as

$$I_{prms} = nI_{Lrms} \tag{3.10}$$

Unlike the half-wave rectifier topology, the output voltage here is a full-wave rectified waveform. This means that the output filter requirement is less stringent as compared to the half-wave rectifier output.

Full-Wave Bridge Rectifier

The full-wave bridge rectifier circuit is shown in Figure 3.5. This is one of the most popular rectifier configurations used in many practical circuits. Here four diodes are used as indicated. When the dot poles of the transformer are positive, then D_1 and D_4 will conduct. When the dot poles are negative, then D_3 and D_2 will conduct thus resulting in a full-wave rectified voltage waveform at the output. It should be noted that the full-wave bridge topology can also be used for direct line rectification wherein one can avoid using the input transformer.

The bridge circuit is normally used for medium-to-high-voltage output applications. Here, the peak-reverse voltage for all the diodes is V_s itself, thus requiring diode with PIV ratings that are half that required for the center-tapped configuration for a given output voltage requirement. However, it should be noted that two diodes at any given instant will conduct and are in series. Hence there will be two diode drops in the bridge rectifier topology as compared to the half-wave or full-wave center-tapped rectifier topologies.

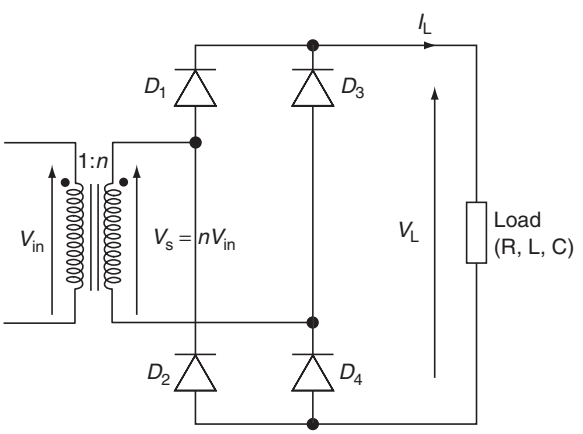


Figure 3.5 | Full-wave bridge rectifier.

In the case of the bridge rectifier circuit also, there is no need for a separate freewheeling diode. This is due to the fact that diode pair D_1, D_2 and diode pair D_3, D_4 act as freewheeling paths for any inductive load. The voltage and current waveforms are shown in Figure 3.6.

Referring to the waveforms shown in Figure 3.6, the load voltage is a pulsating full sinusoid as indicated.

$$V_{\text{Lav}} = \frac{2nV_{\text{m}}}{\pi} \tag{3.11}$$

where V_{Lav} is the average voltage across the load resistor R_{L}; V_{m} is the peak voltage of the input sine wave; n is the transformer turns ratio.

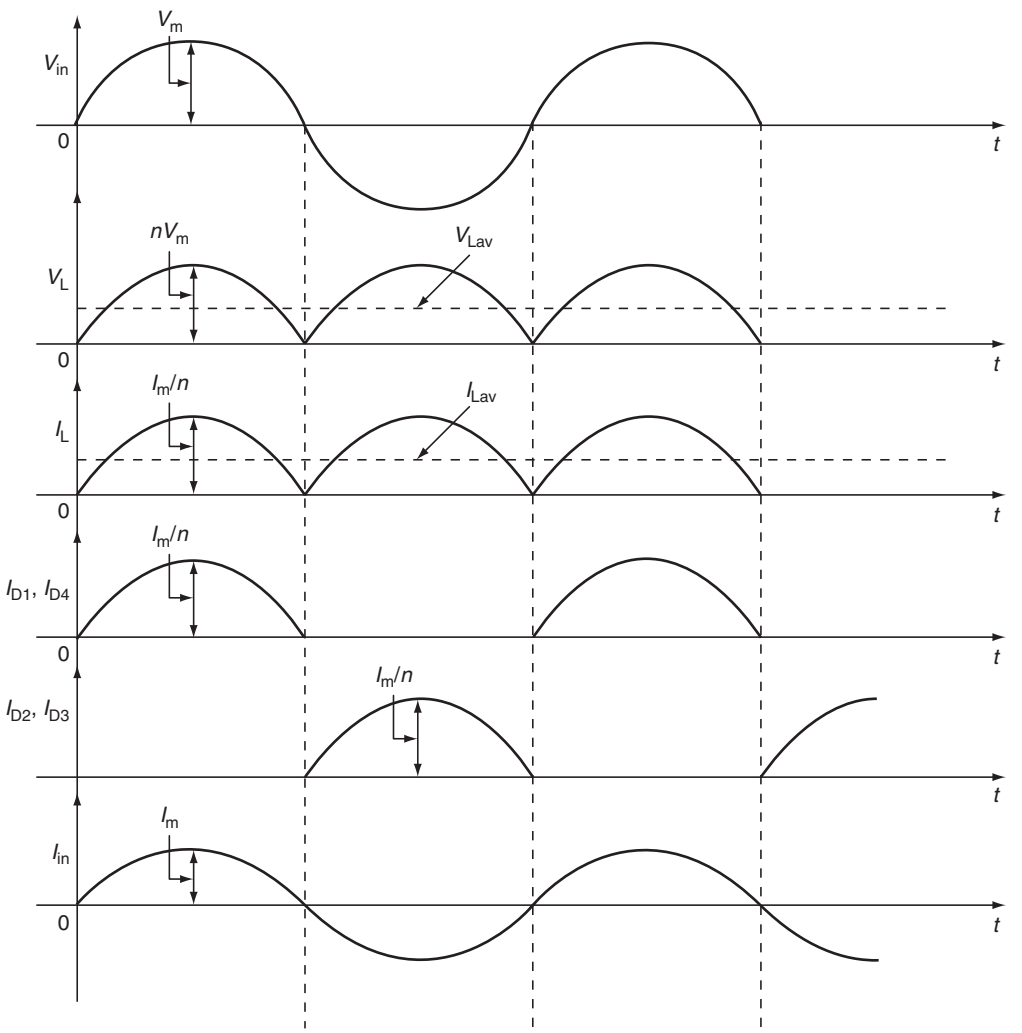


Figure 3.6 | Bridge full-wave rectifier: Waveforms for resistive load.

The average load current is given as

$$I_{\text{Lav}} = \frac{2nV_{\text{m}}}{\pi R_{\text{L}}} \tag{3.12}$$

where I_{Lav} is the average load current. The power delivered to the load is given as

$$\begin{aligned} P_{\text{o}} &= \frac{2}{T}\int_{0}^{T/2} (nV_{\text{m}}\sin\omega t)\left(\frac{nV_{\text{m}}\sin\omega t}{R_{\text{L}}}\right)\text{d}t \\ &= \frac{n^2V_{\text{m}}{}^2}{2R_{\text{L}}} = \left(\frac{nV_{\text{m}}}{\sqrt{2}}\right)\left(\frac{nV_{\text{m}}}{\sqrt{2}R_{\text{L}}}\right) = V_{\text{Lrms}}I_{\text{Lrms}} \end{aligned} \tag{3.13}$$

where V_{Lrms} is the rms voltage across the load resistor and I_{Lrms} is the rms current flowing through the load resistor.

The current through the secondary of the transformer is a full-wave rectified waveform. The rms value of the secondary winding current is given as

$$I_{\text{srms}} = I_{\text{Lrms}} \tag{3.14}$$

The primary or line side current is a full sine waveshape that is the reflected secondary winding current. This is indicated in the I_{in} waveform of Figure 3.6. The rms primary or line side current is given as

$$I_{\text{prms}} = nI_{\text{Lrms}} \tag{3.15}$$

The currents through the diodes are half-wave rectified waveforms as indicated in Figure 3.6. As in the case of the center-tapped full-wave rectifier topology, the output voltage here is a full-wave rectified waveform. This means that the output filter requirement is less stringent as compared to the half-wave rectifier output.

Three-Phase Circuits

Three-phase power is used in medium-to-high-power applications and may be applied to an off-line rectifier, say to produce a nominal 560 V DC bus from a 400 V AC line, or to a transformer whose secondary or secondaries are rectified in various manners. The advantages of using three-phase power as compared to single-phase power are: (a) higher output voltages, (b) lower output ripple, (c) higher input power factor and (d) reduced harmonic distortion of the input current.

Figures 3.7 and 3.8 show the Y–Y and Δ–Y half-wave rectifier circuit configurations. The RYB secondary-side line-to-line voltages, the load voltage and the load currents for a resistive load are shown in Figure 3.9. It can be observed that the load voltage ripple is considerably reduced as compared to that of the single-phase rectifier configurations. Referring to Figure 3.9, it can be observed that the load voltage ripple swing is from the peak voltage nV_{m} to $nV_{\text{m}}\cos(60°)$, that is, $nV_{\text{m}}/2$. Referring to Figure 3.9 and considering the voltage reference axis to be at the peak of the sinusoid, the load voltage contribution from each phase is

$$\frac{1}{2\pi}\int_{-\pi/3}^{\pi/3} nV_{\text{m}}\cos\theta\,\text{d}\theta \tag{3.16}$$

where n is the secondary to primary line-to-line turns ratio.

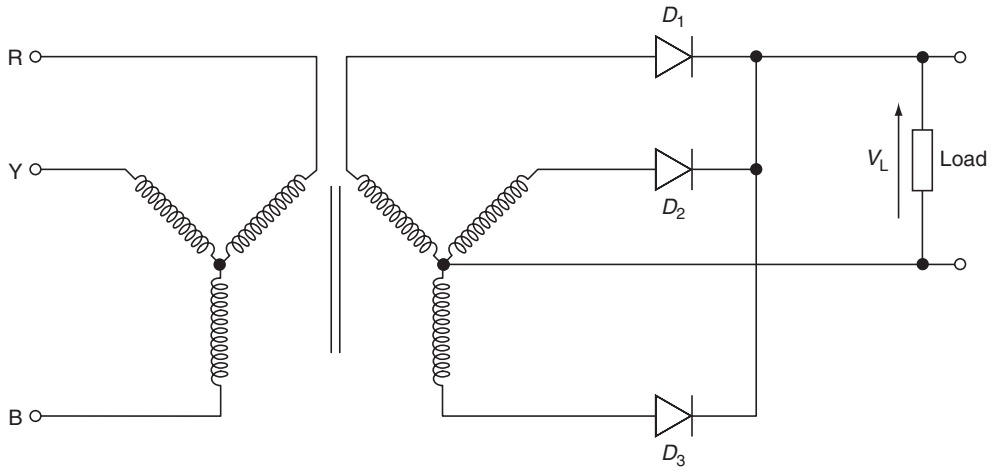


Figure 3.7 | Three-phase half-wave rectifier, Y–Y.

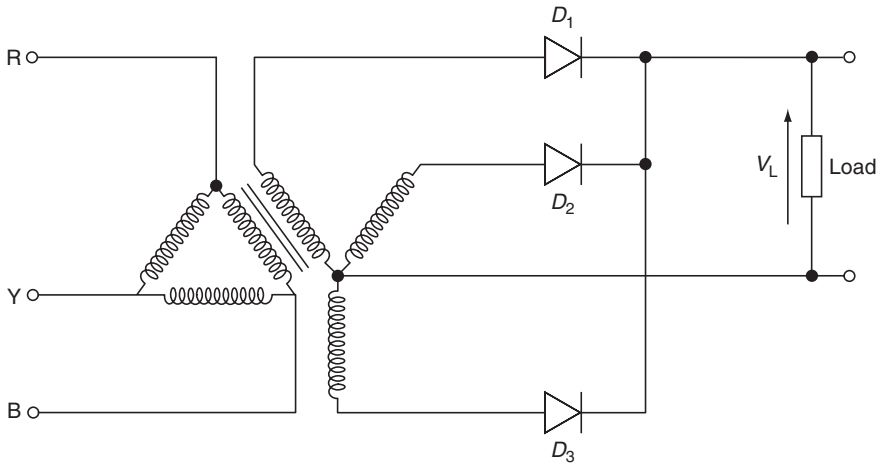


Figure 3.8 | Three-phase half-wave rectifier, Δ–Y.

Each of the three phases contribute to the total load average voltage equally in a period and hence the average load voltage is given as

$$V_{\text{Lav}} = \frac{3}{2\pi}\int_{-\pi/3}^{\pi/3} nV_{\text{m}}\cos\theta \text{d}\theta = \frac{3nV_{\text{m}}}{2\pi}\sin\theta\Big|_{-\pi/3}^{\pi/3} = \frac{3\sqrt{3}nV_{\text{m}}}{2\pi} \tag{3.17}$$

The average load current for a resistive load R_{L} is given as

$$I_{\text{Lav}} = \frac{V_{\text{Lav}}}{R_{\text{L}}} = \frac{3\sqrt{3}nV_{\text{m}}}{2\pi R_{\text{L}}} \tag{3.18}$$

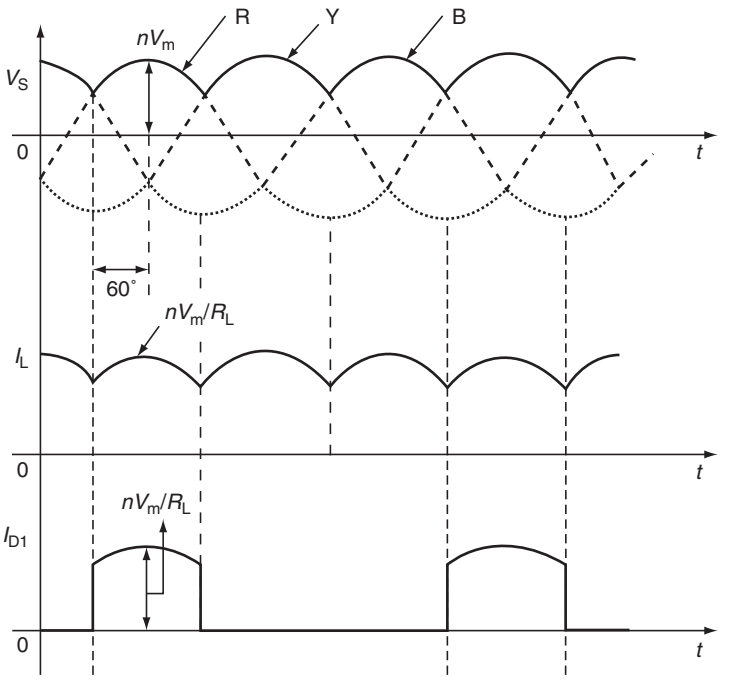


Figure 3.9 Load voltage and current waveforms for half-wave rectifier configurations of Figures 3.7 and 3.8.

The rms value of the load voltage is given as

$$V_{\text{Lrms}} = \sqrt{\frac{1}{2\pi}\int_{-\pi/3}^{\pi/3}(nV_m\cos\theta)^2\,\mathrm{d}\theta\times 3}$$

$$= \sqrt{\frac{3n^2V_m^{\,2}}{2\pi}\int_{-\pi/3}^{\pi/3}\cos^2\theta\,\mathrm{d}\theta} = \sqrt{\frac{3n^2V_m^{\,2}}{4\pi}\left(\theta\Big|_{-\pi/3}^{\pi/3} + \frac{\sin 2\theta}{2}\Big|_{-\pi/3}^{\pi/3}\right)}$$

$$= \sqrt{n^2V_m^{\,2}\left(\frac{1}{2}+\frac{3\sqrt{3}}{8\pi}\right)} = \sqrt{\frac{n^2V_m^{\,2}}{2}\left(1+\frac{3\sqrt{3}}{4\pi}\right)}$$

$$V_{\text{Lrms}} = \frac{nV_m}{\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{4\pi}} \tag{3.19}$$

Similarly, the rms value of the load current for a resistive load is given as

$$I_{\text{Lrms}} = \frac{nV_m}{R_L\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{4\pi}} \tag{3.20}$$

The power delivered to the load is given as

$$P_o = \left[\frac{1}{2\pi}\int_{-\pi/3}^{\pi/3}(nV_m\cos\theta)\cdot\left(\frac{nV_m\cos\theta}{R_L}\right)d\theta\right]\times 3$$

$$= \frac{3n^2V_m^2}{2\pi R_L}\int_{-\pi/3}^{\pi/3}\cos^2\theta\cdot d\theta = \frac{3n^2V_m^2}{4\pi R_L}\int_{-\pi/3}^{\pi/3}(1+\cos 2\theta)\cdot d\theta$$

$$= \frac{3n^2V_m^2}{4\pi R_L}\left(\theta\Big|_{-\pi/3}^{\pi/3} + \frac{\sin 2\theta}{2}\Big|_{-\pi/3}^{\pi/3}\right)$$

$$= \frac{3n^2V_m^2}{4\pi R_L}\left(\frac{2\pi}{3}+\frac{\sqrt{3}}{2}\right)$$

$$P_o = \frac{n^2V_m^2}{R_L}\left(\frac{1}{2}+\frac{3\sqrt{3}}{8\pi}\right) \tag{3.21}$$

Referring to Eqs. (3.19) and (3.20), one can observe that the output power P_o is

$$P_o = V_{Lrms}I_{Lrms} \tag{3.22}$$

Diodes D_1, D_2 and D_3 carry current for only a third portion of each period. The current in diode D_1 is shown in Figure 3.9. The average and the rms values of the diode currents are given as

$$I_{Dav} = \frac{\sqrt{3}nV_m}{2\pi R_L} \tag{3.23}$$

$$I_{Drms} = \frac{nV_m}{R_L\sqrt{2}}\sqrt{\frac{1}{3}+\frac{\sqrt{3}}{4\pi}} \tag{3.24}$$

The secondary windings of the transformer carry the same currents as that flowing through the diodes. Therefore the secondary winding average and rms currents are the same as that of the diodes. The primary winding currents are turns ratio (n) times the secondary currents. In the case of star-connected primary, the winding and line currents are same. In case the primary windings are delta-connected, the line currents are $\sqrt{3}$ times the winding currents.

Figure 3.10 shows a full-wave rectification with a three-phase bridge diode rectifier connected to the secondary as indicated in Figure 3.10. It can be observed that the availability of the neutral point is not essential for rectification in this topology. However, if the neutral point is available then a positive output at the positive output rail with respect to the neutral and a negative output at the negative output rail with respect to the neutral can be obtained in this configuration.

Figure 3.10 shows the Δ–Y full-wave bridge rectifier circuit configurations. The RYB secondary-side line-to-line voltages, the load voltage and the load currents for a resistive load are shown in Figure 3.11. Here the full-bridge topology performs a full-wave rectification. For example, when the secondary-side R phase is positive, diode D_1 conducts and when the R phase goes negative, diode D_4 conducts. In a similar

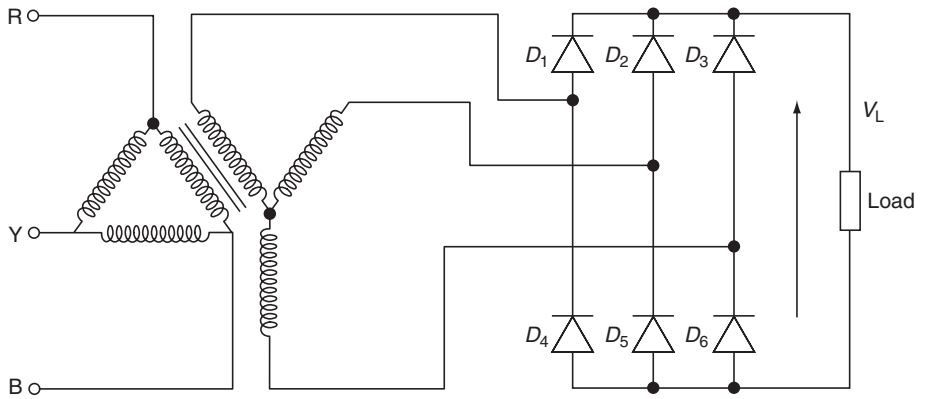


Figure 3.10 | Three-phase full-wave bridge rectifier, Y–Y.

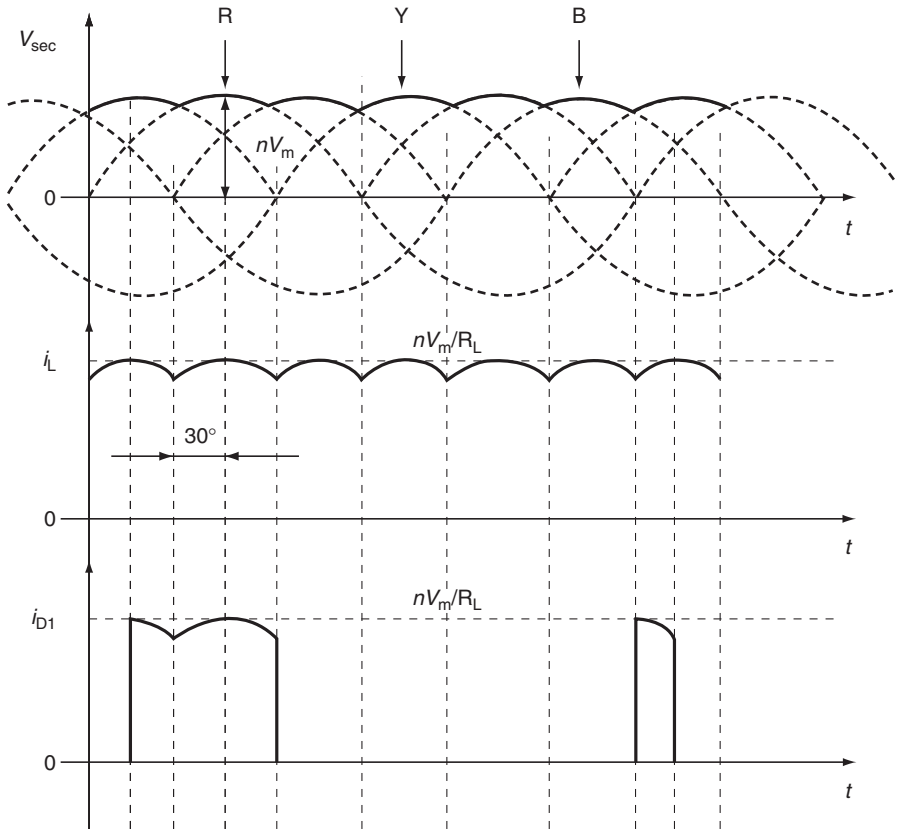


Figure 3.11 | Load voltage and current waveforms for bridge rectifier configurations of Figure 3.10.

manner the diode pairs (D_2, D_5) and (D_3, D_6) bridge arms also operate. It can be noted that the load voltage ripple is further improved as compared to the three-phase half-wave rectifier circuit. Referring to Figure 3.11, it can be observed that the load voltage ripple swing is from the peak voltage nV_m to $nV_m\cos(30°)$, that is, $\sqrt{3}nV_m/2$. There are six ripples in one period. This means that each phase contributes two of the six ripples. Referring to Figure 3.11 and considering the voltage reference axis to be at the peak of the sinusoid, the load voltage contribution from each phase is equal to

$$2\frac{1}{2\pi}\int_{-\pi/6}^{\pi/6} nV_m\cos\theta\,d\theta \tag{3.25}$$

where n is the secondary to primary line-to-line turns ratio.

Each of the three phases contribute to the total load average voltage equally in a period and hence the average load voltage is given as

$$V_{Lav} = 2\frac{3}{2\pi}\int_{-\pi/6}^{\pi/6} nV_m\cos\theta\,d\theta = \frac{3nV_m}{\pi}\sin\theta\Big|_{-\pi/6}^{\pi/6} = \frac{3nV_m}{\pi} \tag{3.26}$$

The average load current for a resistive load R_L is given as

$$I_{Lav} = \frac{V_{Lav}}{R_L} = \frac{3nV_m}{\pi R_L} \tag{3.27}$$

The rms value of the load voltage is given as

$$\begin{aligned} V_{Lrms} &= \sqrt{\frac{1}{2\pi}\int_{-\pi/6}^{\pi/6}(nV_m\cos\theta)^2\,d\theta\times 6} \\ &= \sqrt{\frac{3n^2V_m^{\,2}}{2\pi}\int_{-\pi/6}^{\pi/6}\cos^2\theta\,d\theta} = \sqrt{\frac{3n^2V_m^{\,2}}{2\pi}\left(\theta\Big|_{-\pi/6}^{\pi/6} + \frac{\sin 2\theta}{2}\Big|_{-\pi/6}^{\pi/6}\right)} \\ &= \sqrt{n^2V_m^{\,2}\left(\frac{1}{2}+\frac{3\sqrt{3}}{4\pi}\right)} = \sqrt{\frac{n^2V_m^{\,2}}{2}\left(1+\frac{3\sqrt{3}}{2\pi}\right)} \end{aligned}$$

$$V_{Lrms} = \frac{nV_m}{\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{2\pi}} \tag{3.28}$$

Similarly, the rms value of the load current for a resistive load is given as

$$I_{Lrms} = \frac{nV_m}{R_L\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{2\pi}} \tag{3.29}$$

The power delivered to the load is given as

$$P_o = \left[\frac{1}{2\pi}\int_{-\pi/6}^{\pi/6}(nV_m\cos\theta)\cdot\left(\frac{nV_m\cos\theta}{R_L}\right)d\theta\right]\times 6$$

$$= \frac{3n^2V_m^2}{\pi R_L}\int_{-\pi/6}^{\pi/6} \cos^2\theta \cdot d\theta$$

$$= \frac{3n^2V_m^2}{2\pi R_L}\int_{-\pi/6}^{\pi/6} (1+\cos 2\theta)\cdot d\theta$$

$$= \frac{3n^2V_m^2}{2\pi R_L}\left(\theta\Big|_{-\pi/6}^{\pi/6} + \frac{\sin 2\theta}{2}\Big|_{-\pi/6}^{\pi/6}\right)$$

$$= \frac{3n^2V_m^2}{2\pi R_L}\left(\frac{\pi}{3}+\frac{\sqrt{3}}{2}\right)$$

$$P_o = \frac{n^2V_m^2}{R_L}\left(\frac{1}{2}+\frac{3\sqrt{3}}{4\pi}\right) \tag{3.30}$$

Referring to Eqs. (3.28) and (3.29), one can observe that the output power P_o is

$$P_o = V_{Lrms} I_{Lrms} \tag{3.31}$$

Referring to Figure 3.11, it can be seen from the representative D_1 current waveshape that each diode carries the load current during two ripples in a period or carry an equivalent of one of the six load current ripples in half a period. The D_1 diode current is shown in Figure 3.11. The average and the rms values of the diode currents are given as

$$I_{Dav} = \frac{1}{\pi R_L}\int_{-\pi/6}^{\pi/6} nV_m \cos\theta d\theta = \frac{nV_m}{\pi R_L}\sin\theta\Big|_{-\pi/6}^{\pi/6} = \frac{nV_m}{\pi R_L} \tag{3.32}$$

$$I_{Drms} = \frac{nV_m}{R_L\sqrt{2}}\sqrt{\frac{1}{3}+\frac{\sqrt{3}}{2\pi}} \tag{3.33}$$

3.3 Capacitor Input Filter

The capacitor input filter is an inexpensive and one of the most popular filters that is used for almost all applications and loads that require a DC bus. Capacitor input filters are the most volumetrically efficient means of filtering rectified sine waves and storing energy. The schematic of the capacitor input filter is shown in Figure 3.12. The capacitor charges up to the peak value of the input voltage and tries to maintain this value as the full-wave rectified input drops to zero. The capacitor will discharge through the load until the input full-wave rectified voltage again increases to a value greater than the capacitor voltage. At this point, the diode rectifier will again recharge the capacitor.

The ripple voltage across the filter capacitor is a function of the filter capacitance value, the input frequency and the load current. Considerable importance is given to calculations of ripple amplitude because this parameter influences other design parameters for downstream power conversion devices. In the case of linear power supplies, the minimum capacitor voltage at low line (i.e., minimum input voltage) must be equal to the output voltage plus the minimum voltage which the pass regulator can tolerate while maintaining a regulated output. At high line, the voltage across the pass regulator increases and the regulator must dissipate substantial power.

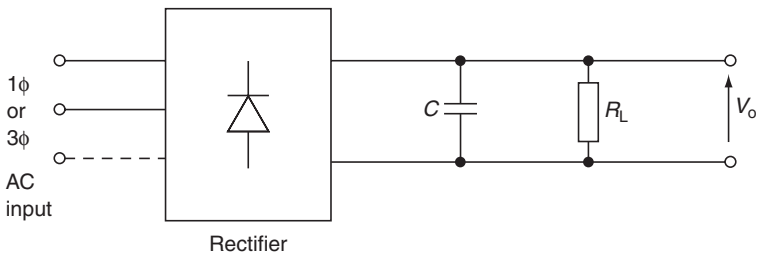


Figure 3.12 | Capacitor input filter.

In switch-mode power supplies, higher ripple voltage may be tolerated since the pulse-width modulator will correct for the DC bus ripple variations without an increase in power dissipation. The filter capacitance may be chosen (for a desired output ripple) or the output ripple may be decided (for a desired capacitance). In many cases, the power supply can operate with a 25% peak-to-peak ripple voltage across the input filter capacitor and a line variation of ±15%.

Design of Capacitor Input Filter Rectifier

Figure 3.13 shows a single-phase full-wave bridge rectifier with capacitor input filter. This is the most common type of AC-to-DC converter used in most applications. The design of the circuit of Figure 3.13 is now discussed in detail.

The input filter capacitor C determines the peak-to-peak ripple voltage component ΔV_r. At low-line voltage, ΔV_r determines the minimum input voltage $V_{m\text{-}min}$ which is an important design consideration and defines the transformer turns ratio required to achieve the specified output voltage at maximum load conditions or maximum duty cycle of any downstream-controlled converter or inverter.

If the input filter capacitor is too small, the resulting large ripple voltage will require the post regulator to have a large line-regulation requirement. This would result in poor transformer utilization and higher ratings for the post regulator power devices. If the input filter capacitor is larger than necessary, it will not only cost more but the recharging capacitive current pulses drawn from the line will also be narrower and

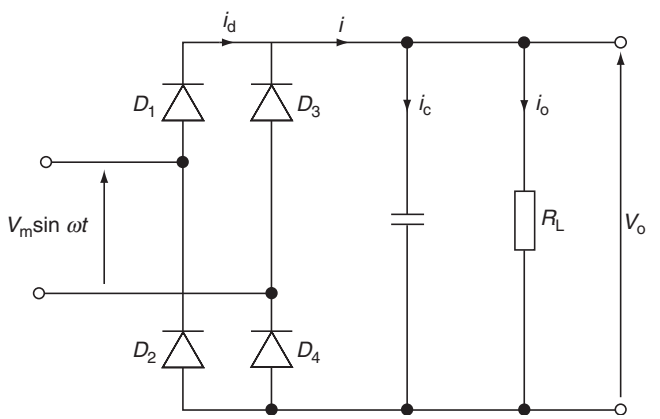


Figure 3.13 | Single-phase rectifier–capacitor input filter.

larger in amplitude. This adversely affects the line power factor and also increases the conducted electromagnetic interference (EMI). The higher rms input line current causes increased losses in the line, diodes and filter capacitor, thus decreasing the efficiency and reliability. Therefore a reasonable rule of thumb is to compromise on a ripple voltage of about 15–30% of the minimum peak line voltage resulting in acceptable capacitor size, weight and cost. It is expected that the downstream converter or inverter will take care of the ripple and line regulation.

Figure 3.14 shows the voltage and current waveforms for the rectifier-filter circuit of Figure 3.13. The capacitor charges only during the period corresponding to the angle δ shown in Figure 3.13. During the remaining period the capacitor is discharging to the load. The current waveforms are approximated to pulsed waveforms as shown in Figure 3.13 without loss of generality from the point of view of design of diodes and capacitor as the current rating obtained for the components would be a conservative value.

During the positive cycle the diode pair (D_1, D_4) conducts and during the negative cycle the pair (D_3, D_2) conducts. The capacitor gets fully charged after four to five cycles. Referring to Figure 3.14, it can be

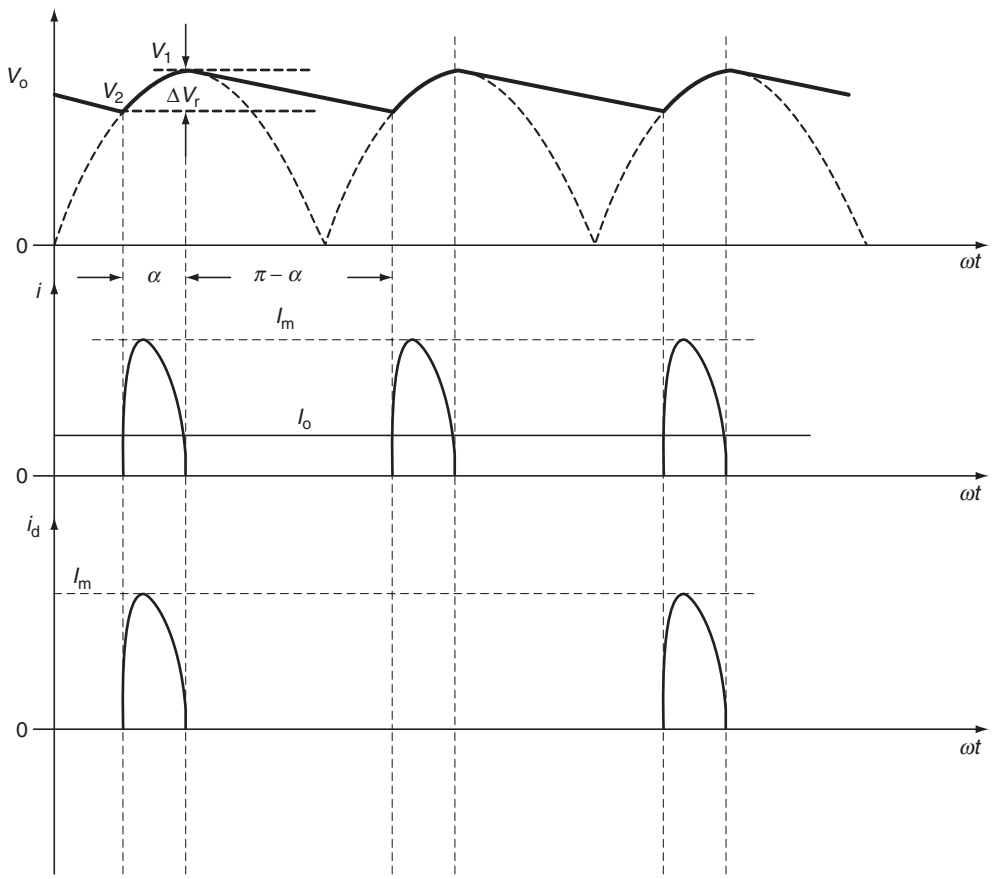


Figure 3.14 | Waveforms of voltage and currents for *C*-filter rectifier as shown in Figure 3.13.

observed that diodes D_1 and D_4 will get forward-biased when the input voltage increases more than V_2. Similarly, diodes D_3 and D_2 get forward-biased when the input voltage goes below V_2 in the negative half-cycle. Only when the rectifier diodes are forward-biased will there be a current flow to charge the capacitor C. The charging up of the capacitor is reflected as an increase in voltage across the capacitor as indicated in the waveform shown in Figure 3.14. Once the capacitor charges to V_1, the input voltage begins to fall wherein the rectifier diodes will become reverse-biased. Now the capacitor discharges into the load with average load current I_o and as a consequence the output voltage V_o will decrease as shown. This process repeats cycle by cycle resulting in the output voltage to have a ripple of ΔV_r. Observe from Figure 3.14 that the capacitor charging current flows only during period α as indicated.

Rectifier Component Selection

In order to select the components of the rectifier, the following parameters must be specified *a priori*.

1. Output voltage, V_o.
2. Output ripple, ΔV_r.
3. Maximum load power, P_o.
4. Nominal input rms voltage, $V_{in,rms} \pm$ % tolerance.
5. Input voltage variation tolerance, %tol.
6. Frequency of input, f_s.

At steady state, the energy gained by the capacitor is equal to the energy lost to the load. In a half-period, the capacitor energy that is given to the load ε_o is

$$\varepsilon_o = \frac{1}{2}C(V_1^2 - V_2^2) \tag{3.34}$$

The output load energy in a half-period of $T/2$ when the capacitor discharges is given as

$$\varepsilon_o = P_o\left(\frac{\pi - \alpha}{\pi}\right)\frac{T}{2} \tag{3.35}$$

From Eqs. (3.34) and (3.35), the output capacitor value is given as

$$C = \frac{\pi - \alpha}{\pi}\frac{2P_o}{(V_1^2 - V_2^2)f_r} \tag{3.36}$$

where f_r is the frequency of the rectified waveform and is equal to $2/T$. Equation (3.36) is the capacitor selection equation. With reference to this equation, the following comments are important. If f_s is the source or line frequency, then

1. for single-phase half-wave rectifiers, $f_r = f_s = 50$ Hz;
2. for single-phase full-wave rectifiers, $f_r = 2f_s = 100$ Hz;
3. for three-phase full-wave rectifiers, $f_r = 6f_s = 300$ Hz.

The output power P_o is a design specification that is determined from the load requirements. V_1 is the peak value of the input voltage. It should correspond to the minimum peak value the input voltage can reach. This is because the capacitor should be selected such that it is capable of providing the specified output energy even under low input voltage situations. Thus,

$$V_1 = V_{m-min} \tag{3.37}$$

where

$$V_{\text{m-min}} = \sqrt{2}V_{\text{in-rms}}\left(1 - \frac{\%\text{tol}}{100}\right)$$

V_2 is given as

$$V_2 = V_1 - \Delta V_{\text{r}} \tag{3.38}$$

where ΔV_{r} is the peak-to-peak ripple voltage and is usually specified at 15–30% of the minimum input voltage.

In Figure 3.14, the currents through the diodes and the capacitors are shown approximated as pulse currents of peak amplitude I_{m} and conducting for a period corresponding to the angle α. As the pulse approximation of the current waveshapes totally enclose the actual current waveshapes any component selection (diode and capacitor) made based on the pulse current waveshapes will definitely work for the actual waveshapes also. It should be noted that approximations based on such engineering judgment will be encountered frequently in circuit design. Such approximations will enable one to obtain close form solutions to many design parameters without compromising on the component ratings.

The peak, average and the rms currents that flow through the diodes and the rms current through the capacitor are required to the calculated so that properly rated diodes and output capacitor can be selected. Referring to Figure 3.14, it can be observed that

$$V_2 = V_1 \cos\alpha$$

and therefore

$$\alpha = \cos^{-1}\frac{V_2}{V_1} \tag{3.39}$$

Let I_{o} be the average value of the load current, then

$$I_{\text{o}} = \frac{P_{\text{o}}}{V_{\text{o-avg}}} \tag{3.40}$$

where $V_{\text{o,avg}} = V_{\text{o-avg}} = (V_1 + V_2)/2$.

Referring to Figure 3.14, the current i flows through the capacitor and also through the load. As the capacitor current has zero average value under steady state, the load current I_{o} is the average value of the current i. Therefore,

$$I_{\text{o}} = I_{\text{m}}\left(\frac{\alpha}{\pi}\right) \tag{3.41}$$

Substitution of Eq. (3.40) into Eq. (3.41) and re-arranging gives

$$I_{\text{m}} = \frac{2\pi P_{\text{o}}}{\alpha(V_1 + V_2)} \tag{3.42}$$

Diode Selection

1. The diode peak current rating should be greater than the value calculated by Eq. (3.42).
2. The rms value of current through diodes is given by $I_{\text{d,rms}} = I_{\text{m}}\sqrt{\alpha/2\pi}$.
3. The average value of current through diodes, $I_{\text{d,avg}} = I_{\text{m}}\alpha/2\pi$.
4. The PIV rating for the diodes should be greater than the maximum peak value that the input voltage will reach. This is given as

$$\text{PIV} > V_{1-\max} = \sqrt{2}V_{\text{in-rms}}\left(1+\frac{\%\text{tol}}{100}\right)$$

5. The average and the rms currents through the diodes are needed to estimate the power dissipated in the diodes. This will reflect in the selection of heat sinks for the diodes.

Capacitor Selection

1. The capacitor value selected should be greater than that calculated by Eq. (3.36). As the capacitor charges up only in one direction, the voltage across the capacitor is unidirectional. Therefore an electrolytic capacitor should be selected.

2. The rms value of current through capacitor is

$$I_{\text{Crms}} = \sqrt{(I_{\text{m}} - I_{\text{o}})^2 \frac{\alpha}{\pi} + I_{\text{o}}^2\left(\frac{\pi-\alpha}{\pi}\right)} \tag{3.43}$$

Equation (3.43) is obtained by referring to Figure 3.14. It can be observed that $(I_{\text{m}} - I_{\text{o}})$ current flows through the capacitor during α out of π periods and I_{o} value of current flows through the capacitor during the remaining $\pi - \alpha$ out of π periods. Squaring the currents and integrating between the angle limits mentioned above and taking the mean over π gives I_{Crms}. The rms value of the capacitor current is used to select the equivalent series resistance (ESR) rating for the capacitor.

3. The voltage rating of the capacitor should be greater than $\text{V}_{\text{1-max}}$.

Remarks on the rectifier–capacitor input filter design:

1. The specifications for the design of the rectifier–capacitor input filter circuit consists of input voltage minimum ($V_{\text{m-min}}$) and maximum ($V_{\text{m-max}}$) limits, desired output voltage peak-to-peak ripple (ΔV_{r}), output power (P_{o}) and the input frequency (f_{s}).

2. The capacitor value is evaluated according to Eq. (3.36). It is essential to take care that the capacitor value is evaluated for the minimum input voltage value.

3. The capacitor is an electrolytic capacitor as the voltage it supports is unidirectional. The voltage rating of the capacitor selected should be greater than $V_{\text{m-max}}$. Either an aluminum electrolytic or a tantalum capacitor is normally chosen.

4. The rms current rating required for the capacitor is calculated using Eq. (3.43). From the manufacturers' datasheets, a capacitor having an rms current rating that is 1.5 to 2 times that calculated is selected.

5. The average, peak and rms current rating of the rectifier diodes are calculated using Eqs. (3.40)–(3.42). From the manufacturers' datasheets, a diode having current ratings that are 1.5 to 2 times the values calculated is selected.

6. The PIV seen by the rectifier diodes is $V_{\text{m-max}}$. Therefore, diodes having PIV ratings that are greater than the value of $V_{\text{m-max}}$ should be selected.

One should note that the design of rectifier-filter circuit with three-phase inputs is similar to that outlined above. In the case of three-phase inputs, the rectifier is a three-phase full-wave bridge rectifier as shown in Figure 3.15. The output V_{o} charges to the peak value of the line-to-line voltage. If V_{m} is the peak value of the line-to-neutral voltage, then V_{o} will charge up to $\sqrt{3}V_{\text{m}}$. The capacitor design equation is same as Eq. (3.36) except that f_{r} is now six times f_{s}. Thus, it is evident that for a given output power and peak-to-peak ripple, the size of the capacitor is smaller with three-phase input source than that for a single-phase input source.

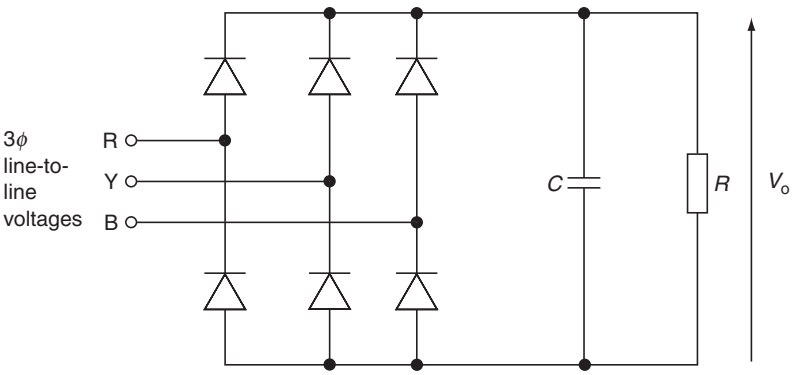


Figure 3.15 | Three-phase capacitor-filter rectifier circuit.

Many products are made for the global markets. The two most common types of mains input that the AC–DC converter may be used with are 115 V and 230 V line inputs. A circuit diagram of a universal or dual input range rectifier filter is shown in Figure 3.16. The charges on the capacitors C_1 and C_2 are balanced by equalizing resistors connected across each capacitor. For 230 V line operation, the switch S is set at position "b". The input rectifiers are now configured as a normal full-wave bridge circuit. For 115 V line operation, the switch S is set at position "a". The input rectifier is now configured as a voltage doubler such that V_o will have the same value as though operating from a 230 V line. While it is technically possible to operate the input section as a bridge at both 230 V and 115 V, the post regulator, which in most cases is a switching regulator, will have to be designed to operate over a much larger input voltage swing which would significantly increase the cost. Therefore, by adopting the above strategy as indicated in Figure 3.16, the post regulator needs to be designed only for 230 V line input.

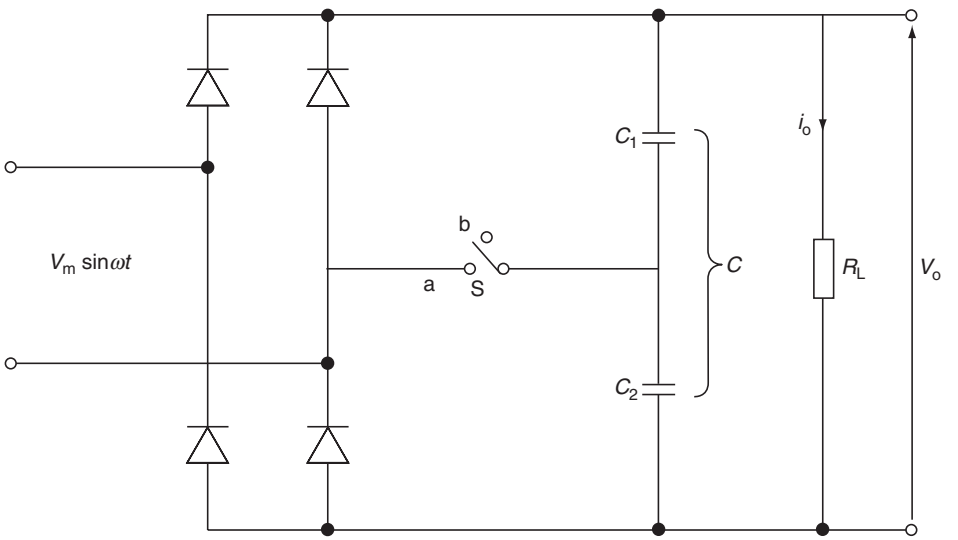


Figure 3.16 | Dual input range rectifier filter.

Hold Time

Let V_h be the minimum voltage required by the load/post regulator to function. This implies that for V_o greater than V_h, the load/post regulator will function as per specification. However, when V_o becomes less than V_h, the load will not get the required minimum voltage and therefore will not become operational. When the input power goes OFF, the time taken by the rectifier output voltage to come down to V_h from the moment the input power went OFF is called the hold time (t_h). The concept of hold time can be visualized from Figure 3.17.

If the rectifier output is powering up a digital controller board or any intelligent microcontroller or microprocessor or digital signal processor board, during the hold time the digital controller will perform safe shut down operations. Therefore, sometimes the output capacitor needs to be calculated based on the minimum hold time requirement of the load. The worst case hold time occurs when the input mains voltage is at a minimum, that is, $V_{m\text{-}min}$ and the power goes OFF when the output capacitor voltage is at V_2 value.

During the hold time, the output capacitor must supply a maximum amount of energy given by

$$\varepsilon_h = P_o t_h \tag{3.44}$$

$$\varepsilon_h = \frac{1}{2} C(V_{m\text{-}min}{}^2 - V_h^2) \tag{3.45}$$

where

$$V_{m\text{-}min} = \sqrt{2} \cdot V_{in\text{-}rms}\left(1 - \frac{\%tol}{100}\right)$$

and V_h is specified as per load requirements. From Eqs. (3.44) and (3.45) the capacitor value for a specified hold time is given as

$$C = \frac{2P_o t_h}{V_{m\text{-}min}^2 - V_h^2} \tag{3.46}$$

The greater of the two values, as calculated from Eqs. (3.36) and (3.46), should be used for selection of the capacitor value such that both the ripple and hold time requirements are met.

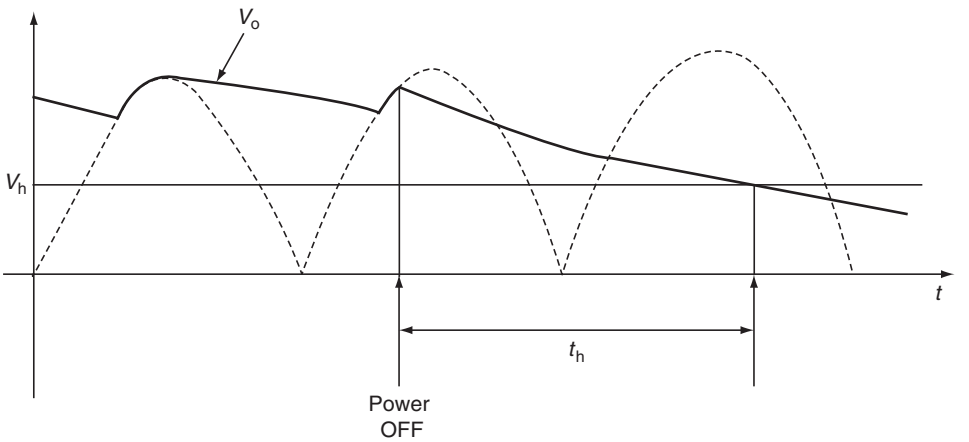


Figure 3.17 | Illustration of hold time.

Turn-ON Currents and Surge Limiting

In a capacitor input filter, the current flow is quite narrow and sharp (as indicated in Figure 3.14). If the AC power is turned ON at or just before the peak of the input voltage waveform, the peak current is limited only by line resistance (R_{line}), ESR of the capacitor(R_{ESR}), rectifier diode dynamic resistance and the input line and/or the leakage reactance of the transformer if present (X_L).

If an input side transformer is used, then the total equivalent resistance of the transformer looking into the secondary is

$$R_{sec} + (R_{line} + R_{pri})\left(\frac{N_s^2}{N_p^2}\right)$$

where R_{sec} and R_{pri} are the winding resistances of the secondary and primary windings of the transformer, respectively; R_{line} is the line or conductor track resistance of the primary side; N_s and N_p are the secondary and the primary number of turns, respectively. The total series resistance R_s is given as

$$R_s = R_{sec} + (R_{line} + R_{pri})\frac{N_s^2}{N_p^2} + 2r_d + R_{ESR} \tag{3.47}$$

where r_d is the dynamic resistance of the diodes. During the capacitor charge duration, two of the four diodes of the full-bridge will conduct. Therefore, two diode dynamic resistances are included in series.

The total series inductance L_s is given as

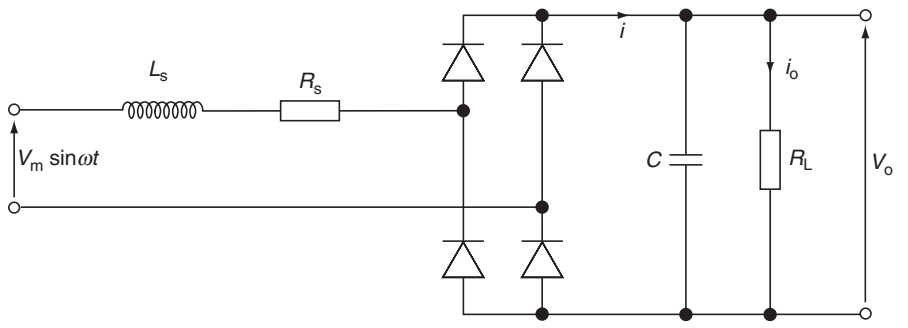
$$L_s = L_{line}\left(\frac{N_s^2}{N_p^2}\right) + L_{leakage} \tag{3.48}$$

L_{line} is the conductor inductance on the primary or the line side and $L_{leakage}$ is the equivalent transformer leakage inductance as seen from the secondary. It should be noted that if the transformer is not present then R_{sec} and R_{pri} values will be zero and N_s/N_p ratio will be unity in Eqs. (3.47) and (3.48). The rectifier–capacitor circuit along with the series impedance is shown in Figure 3.18.

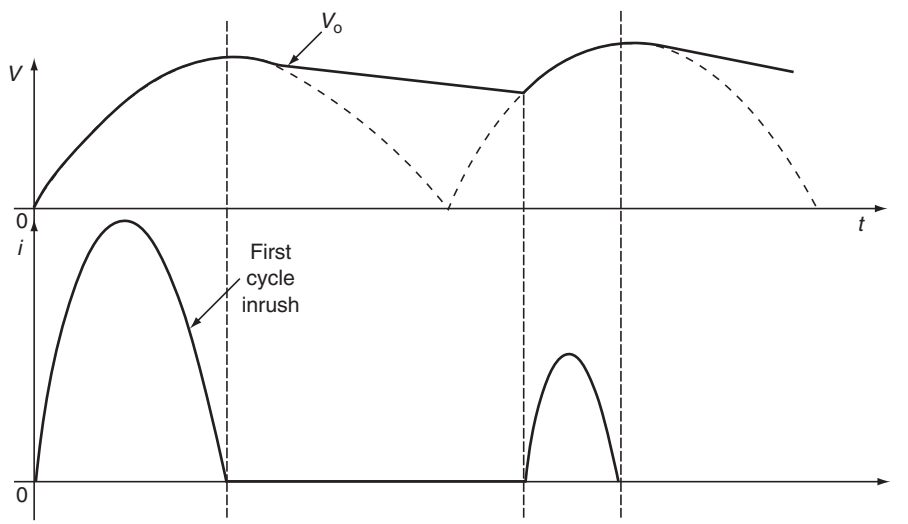
The total source inductance L_s is a function of the inductance in the input line and the leakage reactance of the transformer referred to the secondary. Figure 3.18 shows an off-the-line bridge rectifier with the AC voltage applied when its voltage value corresponds to the maximum. The dynamic equation during the time the capacitor is charging is given as

$$V_{ac} = R_s i + L_s \frac{di}{dt} + \frac{1}{C}\int i\,dt \tag{3.49}$$

From the above equation it may be observed that the startup current i drawn from the source is dependent on the instantaneous amplitude of the input source at startup, the equivalent series circuit resistance R_s, the equivalent series circuit inductance L_s and the output capacitor C. At startup as the capacitor voltage is not yet built up, the current is primarily dependent on the component values of R_s and L_s. In off-the-line bridge rectifier circuits wherein there is no input transformer, the startup current surges could be very high in low L_s circuits. Especially for high-power circuits, the line inductance is very small due to the use of wide conductors. In such cases, the startup surge current will depend primarily on only R_s. This could be detrimental to the components of the rectifier–capacitor circuit. However, in many cases, the inductance of an EMI filter at the input provides sufficient surge protection for the off-the-line rectifier circuit. The startup peak



(a)



(b)

Figure 3.18 (a) Rectifier filter with source inductance and resistance; (b) inrush current in the first cycle.

current may be reduced by a factor of 10 if the power supply is turned ON at the zero crossing of the input wave, as compared to turning ON the power supply at the peak of the input wave.

The substantial startup current surge may be more economical to limit than choosing higher current diodes. Figure 3.19 shows few methods of limiting the startup surge current. In the step-start circuit shown in Figure 3.19(a), R_s is in series with R_1 through which current flows when the input switch is closed, thereby limiting the inrush current. The energizing time of contactor, K_1, is typically from one-to-three cycles of the input frequency which allows the filter capacitor to become charged, after which R_1 is shorted by K_1 contacts to eliminate power dissipation and voltage drop due to inclusion of R_1.

A negative temperature co-efficient power thermistor, as shown in Figure 3.19(b), can also be used to limit the inrush currents. At turn-ON, the "several ohms" resistance in TH_1 limits the inrush current which, in turn, dissipates power in the thermistor. This power dissipation raises the temperature of the device and the resistance drops to a low value for normal operation. However, the thermal time constant of the thermistor must be considered. If a power supply has been operating at near no load for some time, the resistance of TH_1 will be higher than at full load. When full load is applied, the voltage drop across TH_1 will cause the output to drop substantially or in the case of a regulated supply, the output may go out of regulation. If the power supply is operating at full load where TH_1 is very low in resistance and a short power interruption occurs (long enough for the input filter capacitor to discharge), then a high inrush current will occur when the mains voltage returns. This is because the thermistor resistance has not yet recovered to its high cold state value.

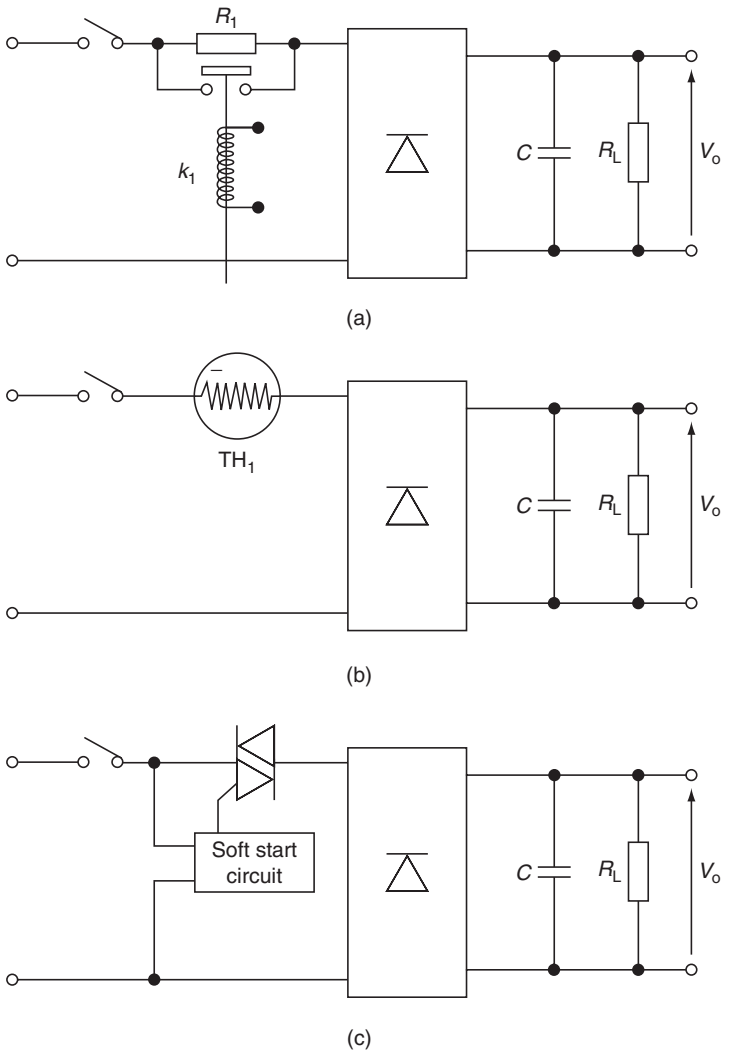
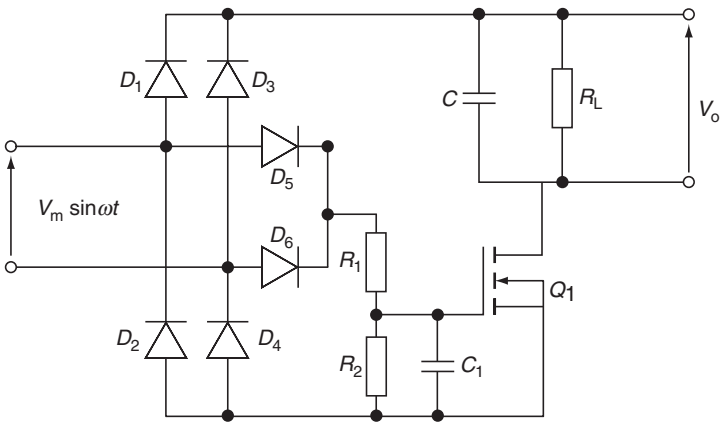


Figure 3.19 | Surge current-limiting circuits.



(d)

Figure 3.19 | *Continued*.

The circuit shown in Figure 3.19(c) utilizes a TRIAC Q_1. The soft-start control circuit is phase controlled at turn-ON to limit inrush current. Initially, the conduction period of each half-cycle is very short, but then increases with time until full conduction is reached. Three precautions should be observed with this circuit. First, the conduction period should increase linearly every cycle at turn-ON to prevent a DC component to be applied to the transformer primary if a transformer is used. In such cases, it may be better to locate the TRIAC on the secondary side. Second, the control circuit should have mains-interruption sensing, with a time constant sufficient to permit soft-start operation when the mains again returns; otherwise, a high inrush current will occur. Third, the control circuit should provide continuous turn-ON signal to Q_1 during desired full conduction. Otherwise, Q_1 will turn-OFF when the input current, which is discontinuous, goes to zero.

An input rectifier and capacitor filter is shown in Figure 3.19(d) with an N-channel power MOSFET inserted in the DC return line. Diodes D_1–D_4 rectify the input to charge C and supply power to R_L when Q_1 is ON. The MOSFET has two main advantages over a bipolar transistor. First, the high input impedance of the gate allows high-resistance values for the biasing resistor, thus minimizing power dissipation. Second, the MOSFET is immune to forward-biased secondary breakdown, although the safe operating area (SOA) must still be observed. At turn-ON, C_1 charges exponentially through R_1 to a final steady-state value determined by $R_2/(R_1 + R_2)$. C_1 charges to a few volts before the gate threshold voltage is reached, at which time Q_1 begins to conduct in the active region and act like a resistor. C reaches full charge in a few cycles, determined by the time constant of $R_1R_2C_1/(R_1 + R_2)$. The addition of diodes D_5 and D_6 serves to decouple C_1 from C during operation and allows C_1 to discharge through R_2 when input power is removed. If D_5 and D_6 were omitted and R_1 were connected to C, the voltage across C (as C discharges) would maintain a charge on C_1 even though input power is momentarily removed. Thus, when input power is again applied, the charge on C_1 could be sufficient to bias Q_1 into full conduction. Complete discharge of C_1 during power interruptions could be achieved by paralleling C_1 with a PNP transistor such that the base is biased by a voltage divider from D_5, D_6 to DC return. With the power ON, the transistor would be reverse-biased but with power OFF, the transistor would conduct to rapidly discharge C_1. Inserting a small resistor between the source terminal of Q_1 and the negative terminal of C_1 can act as a current limiter. As source current increases, V_{gs} decreases thus increasing V_{ds} thereby limiting the inrush current.

3.4 Power Factor

Power factor (*pf*) gives the quality measure of a circuit. If the *pf* is unity, the input power drawn from the source will be the load power and the power that gets dissipated in the various components. However, if the *pf* is low, the input source should be rated for a much higher power than the required load power. This would imply higher line losses resulting in lower efficiencies. Referring to Figure 3.14, the input voltage waveform is sinusoidal, but the current is not. It should be noted that the average current required by the load is I_o but the peak current demand from the source is I_m which is much higher than I_o. This implies a low *pf* for the rectifier–capacitor circuit. Further, as the line current flows only for a period corresponding to α, there is a voltage drop in the line impedance only during this period. This leads to distortion in the input applied voltage. For such cases, what is the *pf*? And how is it computed and measured?

Consider a voltage source of an arbitrary waveshape that is periodic with period T. It is connected to a pure resistive load R as shown in Figure 3.20. The current waveshape will exactly resemble the voltage waveshape. The entire power drawn from the source is given to the load R. This means that no power is used for kinetic storage (in inductor) or potential storage (in capacitor) or returned back to the input. The power that is given to load R is called the active power and the power that is either stored or returned back to the input is called the reactive power. In the case of the circuit as described above the reactive power is zero. The power that is given to the resistive load, P_{res}, is given by

$$P_{res} = \frac{1}{T}\int_0^T v \times i \times \mathrm{d}t \tag{3.50}$$

where v and i are the instantaneous values of the current and voltage across the load resistor R.

If the input source is a voltage source, the reference waveshape for the source currents is that corresponding to the voltage source, and if the input source is a current source, the reference waveshape for the source voltage is that corresponding to the current source. For now, consider that the input source is a voltage source and its waveshape is defined. Let this voltage source of defined waveshape be connected to an arbitrary load. The power delivered from the source is given as

$$P_{load} = \frac{1}{T}\int_0^T v_i \times i_L \times \mathrm{d}t \tag{3.51}$$

where v_i is the instantaneous value of input source voltage that is across the arbitrary load; i_L is the instantaneous value of the current flowing through the source for the specified arbitrary load.

Measure or estimate the peak current I_m that flows from the source for the specified load. Replace the load with a resistive load R_{ref} that has a value which results in the same peak current through the source and load. For a resistive load, the peak current occurs at the peak of the input voltage waveshape, V_m. This would

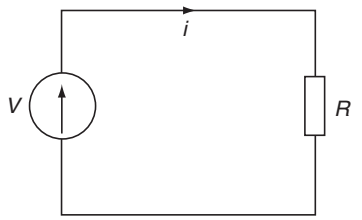


Figure 3.20 A voltage source connected to a resistive load.

be the reference load for the input voltage source for a specified peak current I_m. Thus the value of the reference load resistor R that should be selected for I_m is

$$R_{ref} = \frac{V_m}{I_m} \tag{3.52}$$

where V_m is the peak of the input source voltage; I_m is the measured or estimated peak current flowing through the source for the specified arbitrary load.

The power that is delivered from the input source with the reference resistive load as determined above is given as

$$P_{ref} = \frac{1}{T}\int_0^T v_i \times i_R \times dt \tag{3.53}$$

where P_{ref} is the reference power delivered from the input source with the reference load; v_i is the instantaneous value of input source voltage that is across the arbitrary load; i_R is the instantaneous value of the current flowing through the source for the reference resistive load.

The *pf* is defined as the ratio of the power delivered to the arbitrary load to the reference power delivered to the reference resistive load wherein the same input peak current is maintained. Thus *pf* is given by

$$pf = \frac{P_{load}}{P_{ref}} = \frac{\frac{1}{T}\int_0^T v_i \times i_L \times dt}{\frac{1}{T}\int_0^T v_i \times i_R \times dt} = \frac{\int_0^T v_i \times i_L \times dt}{\int_0^T v_i \times i_R \times dt} \tag{3.54}$$

The *pf* gives a measure of the departure of the arbitrary load from the resistive load. If *pf* is unity, the arbitrary load is equivalent to the resistive load. If *pf* is less than unity, it implies that there are either kinetic or potential energy storage components in the arbitrary load.

EXAMPLE 3.1

RLC Load

Consider a series RLC load connected to a sinusoidal voltage source as shown in Figure 3.21.

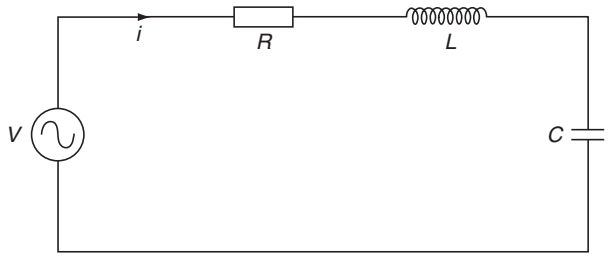


Figure 3.21 Series RLC load.

The current waveform leads or lags the voltage waveform by an angle θ. One can apply Eq. (3.54) to obtain the *pf* for this RLC load supplied from a sinusoidal voltage source. Here

$$v_{\mathrm{i}} = V_{\mathrm{m}} \sin(\omega t)$$

$$i_{\mathrm{L}} = I_{\mathrm{m}} \sin(\omega t - \theta)$$

where

$$\theta = \tan^{-1}\left(\frac{\omega L}{R}\right) - \tan^{-1}\left(\frac{1}{\omega CR}\right)$$

The reference load resistor R_{ref} is given by

$$R_{\mathrm{ref}} = \frac{V_{\mathrm{m}}}{I_{\mathrm{m}}}$$

The *pf* for the series RLC load can be estimated using Eq. (3.54). This is given by

$$pf = \frac{\int_0^T V_{\mathrm{m}} \sin(\omega t) I_{\mathrm{m}} \sin(\omega t - \theta)\mathrm{d}t}{\int_0^T V_{\mathrm{m}} \sin(\omega t) V_{\mathrm{m}} \sin(\omega t)/R_{\mathrm{ref}}]\mathrm{d}t} = \frac{V_{\mathrm{m}} I_{\mathrm{m}} \int_0^T \sin(\omega t)\sin(\omega t - \theta)\mathrm{d}t}{V_{\mathrm{m}} I_{\mathrm{m}} \int_0^T \sin(\omega t)\sin(\omega t)\mathrm{d}t} = \cos(\theta)$$

EXAMPLE 3.2

Rectifier–Capacitor Filter Load

Consider the rectifier–capacitor filter load that is supplied by the sinusoidal voltage source. Here

$$v_{\mathrm{i}} = V_{\mathrm{m}} \sin(\omega t)$$

i_{L} is the discontinuous waveform i as shown in Figure 3.14. This current flows every half-cycle for a period corresponding to α. The peak value I_{m} of this current waveform can be measured using an oscilloscope.

From the estimate of the peak current that flows through the input source, the reference load resistor R_{ref} is given by

$$R_{\mathrm{ref}} = \frac{V_{\mathrm{m}}}{I_{\mathrm{m}}}$$

and then

$$i_{\mathrm{R}} = I_{\mathrm{m}} \sin(\omega t)$$

From Eq. (3.54), the *pf* for the rectifier–capacitor filter load can be obtained as follows:

$$pf = \frac{\frac{1}{T}\int_0^T V_{\mathrm{m}} \sin(\omega t) i \mathrm{d}t}{\frac{1}{T}\int_0^T V_{\mathrm{m}} \sin(\omega t) I_{\mathrm{m}} \sin(\omega t)\mathrm{d}t} = \frac{P_{\mathrm{o}} + P_{\mathrm{loss}}}{V_{\mathrm{m}} I_{\mathrm{m}}/2} = \frac{2(P_{\mathrm{o}} + P_{\mathrm{loss}})}{V_{\mathrm{m}} I_{\mathrm{m}}} \tag{3.55}$$

where P_{loss} is the consolidated losses of the diodes, ESR of capacitor and the transformer losses if transformer is present. In higher power circuits where the losses are negligible compared to the output power, the *pf* for the rectifier–capacitor filter circuit reduces to

$$pf = \frac{2P_o}{V_m I_m} \tag{3.56}$$

The value of I_m has to be measured. However, one can obtain an approximate estimate of the *pf* by using the relationship for I_m given by Eq. (3.42) in Eq. (3.56):

$$pf = \frac{\alpha(V_1 + V_2)}{\pi V_m}$$

where

$$\alpha = \cos^{-1}\left(\frac{V_2}{V_1}\right)$$

Observe from the above representation that as the output ripple decreases, the time duration for which the diodes conduct, α, will reduce. This implies that for the same output power, the current peak I_m will be larger, resulting in *lower pf*. The *pf* can be improved by

1. introducing an inductor in series;
2. increasing the diode conduction angle, by-means of controlled switching of power switches. This is called the unity *pf* converter.

The former method will be discussed in the next section. However, the latter method will be discussed in Chapter 10 after discussion of the control principles.

3.5 Rectifier–LC Filter

The rectifier–LC filter consists of an inductor that is placed between the rectifier and the capacitor filter. Since current cannot change instantaneously in the inductor, inrush currents at turn-ON are reduced. Further, as the inductor has a smoothening effect on the current, the output voltage ripple is reduced as compared to the rectifier–capacitor filter. Most importantly, introduction of the inductor widens the conduction angle of the diodes, thereby improving the *pf* significantly. Alternately one may argue that the capacitor is a potential energy storage element and the inductor being a kinetic energy storage element offsets the effect of the capacitor such that the input source sees an impedance that is closer to the resistive load.

However, the major drawback is the inductor itself. The inductor will have to be designed for the supply frequency of 50 Hz. This will make the size and cost of the inductor prohibitively large. This significant drawback has prevented this topology from becoming popular in commercial equipments.

A single-phase full-wave rectifier with LC filter is shown in Figure 3.22(a). During the peak portions of the voltage waveform, energy is stored in the inductor and during the valley portion of the voltage waveform, it is transferred to the capacitor and load. The waveforms for the rectifier–LC filter are shown in Figure 3.22(b).

Referring to Figure 3.22(b), the rectified waveform contains an average voltage component and an AC component. As the rectified waveform has even wave symmetry, it can be deduced that the AC component contains only even harmonics. The Fourier series of the rectified waveform is given as

$$V_{rect} = \frac{2V_m}{\pi}\left(1 - \frac{2}{3}\cos 2\omega t - \frac{2}{15}\cos 4\omega t - \frac{2}{35}\cos 6\omega t - \ldots\right) \tag{3.57}$$

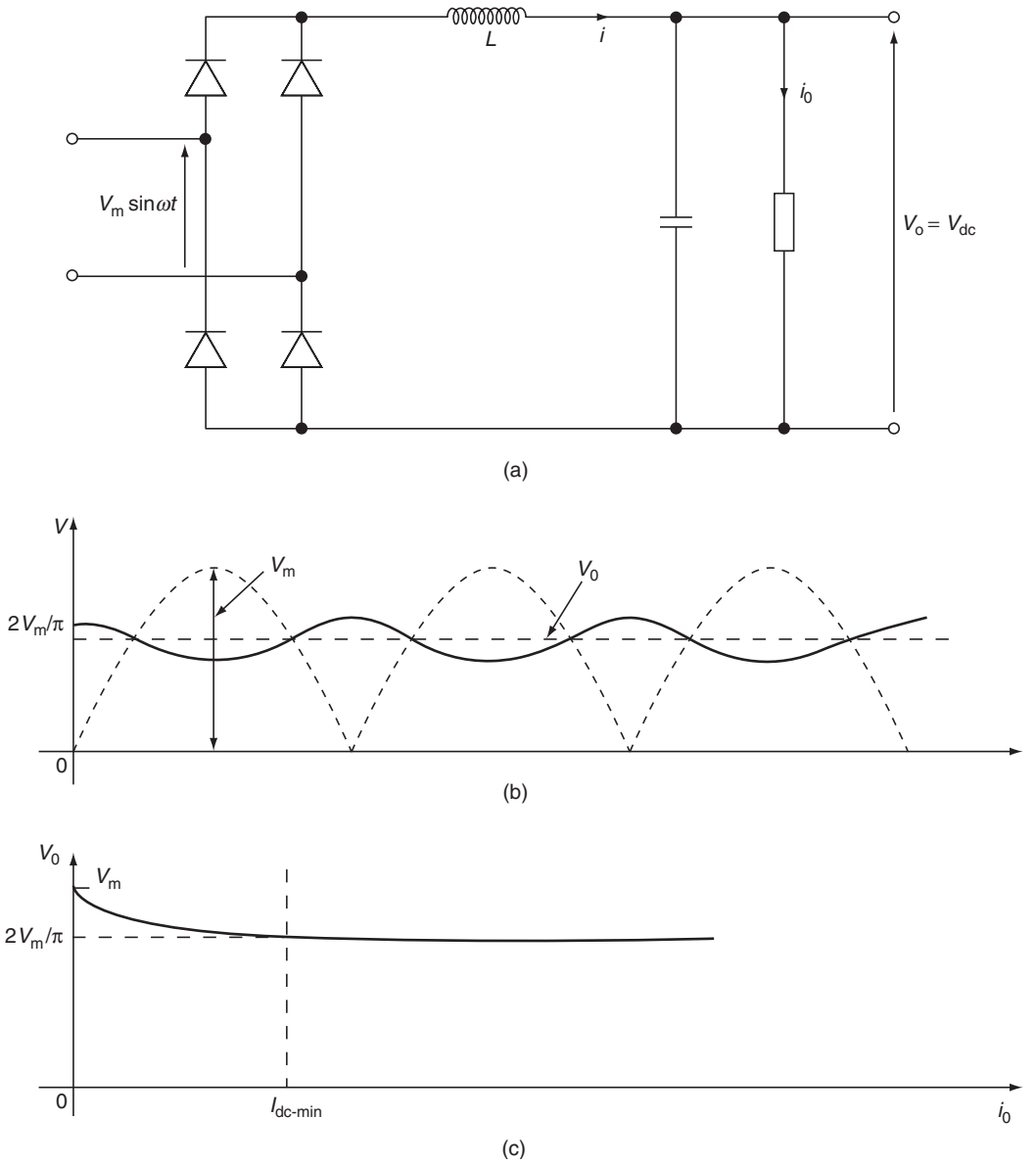


Figure 3.22 | (a) Rectifier–LC filter circuit; (b) voltage waveform; (c) output voltage versus load.

where V_m is the peak value of the rectified sinusoidal voltage; $\omega = 2\pi f$ is the fundamental radian frequency and f is the line frequency.

The rectified waveform Fourier component representation given in Eq. (3.57) can be re-written as

$$V_{rect} = V_{DC\text{-}part} + V_{AC\text{-}part} \tag{3.58}$$

where

$$V_{\text{DC-part}} = \frac{2V_{\text{m}}}{\pi} \tag{3.59}$$

$$V_{\text{AC-part}} = \frac{2V_{\text{m}}}{\pi}\left[\sum_{n=2}^{\infty}\left(\frac{-2}{n^2-1}\right)\cos n\omega t\right] \tag{3.60}$$

where n is the even harmonic number.

The LC-filter circuit is an averaging circuit or a low-pass circuit. It will remove the AC components and pass only the DC part. Therefore, the output voltage V_{o} primarily contains only the average or DC part of the rectified waveform.

$$V_{\text{DC}} = V_{\text{DC-part}} = \frac{2V_{\text{m}}}{\pi} \tag{3.61}$$

The AC part of the rectified waveform gets attenuated due to the capacitive reactance. Thus the AC part of the rectifier waveform that appears at the output is given as

$$V_{\text{ripple}} = V_{\text{AC-part}}\left(\frac{X_{\text{C}} \| R_{\text{L}}}{X_{\text{L}} + X_{\text{C}} \| R_{\text{L}}}\right) \tag{3.62}$$

Thus

$$V_{\text{o}} = V_{\text{DC}} + V_{\text{ripple}}$$

For a good LC-filter circuit, $X_{\text{C}} << R_{\text{L}}$ and $X_{\text{C}} << X_{\text{L}}$. Therefore, the output ripple voltage given in Eq. (3.62) can be approximated as

$$V_{\text{ripple}} = V_{\text{AC-part}}\left(\frac{X_{\text{C}}}{X_{\text{L}}}\right) \tag{3.63}$$

The discussion above assumes that the current through the inductor is continuous. The desired output voltage ripple is achievable only if the inductor current is continuous. As the load resistance R_{L} increases, the DC or the average part of the current through the inductor that flows through R_{L} reduces. As R_{L} is further increased towards open circuit, the inductor current gradually reduces and becomes discontinuous. When this happens, the output voltage will tend to rise toward the peak input voltage, V_{m}, as shown in Figure 3.22(c). Current $I_{\text{dc-min}}$ is the load value at which the inductor current becomes discontinuous and each rectifier diode conducts for less than 180°. The load $I_{\text{dc-min}}$ is the minimum load current for which the inductor is in continuous conduction.

If the minimum load is specified, then the inductor value can be designed such that the inductor current is continuous even at minimum load and the output voltage is as given by Eq. (3.61). The inductor has a DC component and an AC component of current. For the current to be continuous, the DC component $2V_{\text{m}}/\pi R_{\text{L-max}}$ must be equal to or greater than the AC component. As the capacitive reactance $X_{\text{C}} << X_{\text{L}}$, the AC component of the inductor current that comprises predominantly of the second-harmonic component is $4V_{\text{m}}/3\pi X_{\text{L}}$. Thus

$$\frac{2V_{\text{m}}}{\pi R_{\text{L}}} \geq \frac{4V_{\text{m}}}{3\pi X_{\text{L}}} \tag{3.64}$$

where $X_L = 2\omega L$. From the above inequality, the following inductor value selection criterion is obtained.

$$L \geq \frac{R_{L\text{-max}}}{3\omega} \tag{3.65}$$

where $R_{L\text{-max}}$ corresponds to the load resistance value at minimum load. The input source considered for the LC-filter circuit is a single-phase sinusoidal source. On rectification, the number of rectified pulses per period is two. If a three-phase source is rectified then the number of rectified pulses per period is six. If such a rectified output is fed to the LC filter, then the general equation for inductance value of polyphase rectifiers can be obtained along similar lines as discussed and is given as

$$L \geq \frac{2R_{L\text{-max}}}{p(p^2-1)\omega} \tag{3.66}$$

where p is the number of rectified pulses per fundamental line period.

From Eq. (3.63) it can be observed that a larger L will result in a smaller output voltage ripple. From Eq. (3.66) it can be noted that when the load is light, R_L is large and a large value of L is needed if the ripple is not to be excessive. This means that the series inductance should have a large value at no load and may be allowed to decrease as the load increases. Reactors can be designed wherein the inductance decreases as the DC current through inductor increases by making use of the non-linear *B–H* magnetic characteristic of the inductor core material. This can be a solution to the bulky and expensive constant inductance chokes that need more iron to avoid saturation at larger currents.

Alternately, a bleeder resistor could be added across the output filter capacitor such that the minimum load is guaranteed. In this case, the power dissipation and rating of the bleeder resistor and reduction in overall efficiency must be considered. The value of the bleeder resistor for a single-phase full-wave rectifier is

$$R_B = \frac{V_{DC}}{I_{dc\text{-min}}} \tag{3.67}$$

Referring to Figure 3.22(c), the current $I_{dc\text{-min}}$ is the value of the load current at which the current in the inductor becomes discontinuous or in other words the DC component of the inductor current is equal to the peak of the AC component of the inductor current. At still lighter loads, the DC component becomes lesser than the peak AC component and leads to discontinuous currents in the inductor. Therefore, by introducing a bleeder resistor R_B across the output capacitor C the rectifier filter will always see a minimum load that is just enough to make the current in the inductor continuous. From Eq. (3.65) it can be argued that the bleeder resistor that will provide the minimum load for continuous conduction of the inductor current is

$$R_B = 3\omega L \tag{3.68}$$

and for polyphase input sources, the general equation for the bleeder resistor is

$$R_B = \frac{p(p^2-1)\omega L}{2} \tag{3.69}$$

where p is the number of rectified pulses per fundamental line period.

Output Ripple

Equation (3.63) gives the amount of ripple voltage in the output. The inductor has a high reactance and the capacitor has a low reactance at the harmonic frequencies and the following discussion considers only

the predominant second-harmonic effect, with negligible error in analysis. As the second harmonic is the dominant effect with respect to the ripple voltage, the ripple waveshape can be considered to be sinusoidal. Referring to Eqs. (3.60) and (3.63), the rms value of the output ripple voltage is given as

$$V_{\text{ripple-rms}} = \frac{4V_{\text{m}}}{3\sqrt{2}\pi}\frac{X_{\text{C}}}{X_{\text{I}}} = \frac{V_{\text{m}}}{3\sqrt{2}\pi\omega^2 LC} \tag{3.70}$$

The preceding equation provides the *LC* product, but individual component values must still be determined. For a specified value of the minimum load, the value of *L* can be calculated. From the *LC* product, the value of *C* can then be estimated.

Turn-ON Current

An important consideration with inductor input filters is the frequently under-damped nature of the circuit at turn-ON. This is especially true of soft-start switch-mode power supplies because the input filter capacitor will charge to a voltage higher than the peak input voltage and the "downstream" switching transistors will experience a higher-than-normal voltage when switching commences. This effect has to be taken into account while designing post regulators.

Design Summary

In the design of the rectifier–LC filter circuit, the input voltage, the minimum load requirements and the output rms ripple are specified. Based on these inputs:

1. The *LC* value is evaluated using Eq. (3.70).
2. The value of *L* for minimum load is calculated using Eq. (3.65).
3. The value of *C* is calculated from *LC* product and the value of *L*. Ensure that $X_{\text{C}} << X_{\text{L}}$ at the second harmonic.
4. The rectifier diodes should be rated to carry the inductor current. Considering the second harmonic as the dominant effect, this is given as

$$I_{\text{d}} = \frac{2V_{\text{m}}}{\pi R_{\text{L-min}}} + \frac{4V_{\text{m}}}{6\pi\omega L}$$

5. The PIV ratings for the diodes should be greater than V_{m}.

3.6 Controlled Rectifiers

In the previous sections, it is seen that the rectified DC output voltage is dependent on the amplitude of the AC input voltage and the load. The output voltage is therefore unregulated. As the input voltage varies, so does the output voltage. The output voltage increases as the load approaches an open circuit. This section presents phase-control topologies that may be employed to provide a regulated output. Line and load regulation are achieved by using thyristors family of power switches to control the conduction angle of the sine wave input in response to some desired control. TRIACs are used to control AC load voltages, as in light dimmer assemblies. For high-power AC loads, parallel back-to-back SCRs can be used. Power supplies utilize rectifiers and TRIACs or SCRs as the control element to provide a desired DC output from single-phase or three-phase inputs.

Single-Phase Power Circuits

Few topologies of single-phase power circuits are shown in Figure 3.23. The full-wave center-tap with resistive load is used primarily for low-voltage outputs, as there is only one forward SCR drop. The resistive load of Figure 3.23(a) is replaced by an inductor filter in Figure 3.23(b). In this case, a freewheeling diode D is inserted to provide a path for the inductor current when the SCRs turn OFF. The bridge circuits of Figures 3.23(c)–(f) are used primarily for higher output voltage applications. Figures 3.23(a)–(d) have the advantage of common cathodes that permit firing pulse to be applied to both SCRs without need for isolation between the two SCR gate drive circuits. They also permit simultaneous application of trigger pulses from a pulse transformer with a single secondary if isolation is required. However, individual gate resistors to each SCR are recommended. The anode terminal is normally the case of the SCR in stud-mounted devices. The circuit of Figure 3.23(e) allows the SCRs to be mounted to a common heat sink without isolation, while the diodes (cathode is normally the case of the device) are mounted to another heat sink without isolation. The two SCRs are connected in series in Figure 3.23(f), which has the advantage of eliminating the freewheeling diode since this function is performed by the two bridge diodes in series. However, isolated trigger pulses to the SCRs must be provided for the circuits in Figures 3.23(e) and (f). In Figure 3.23(g), four SCRs are utilized. Here, output polarity reversal is possible.

Resistive Load

Figure 3.24 shows the output voltage waveforms at various firing angles for resistive loads in single-phase circuits given in Figure 3.23. Referring to Figure 3.24, the average value or the DC value of the rectified waveform, triggered at firing angle α is given as

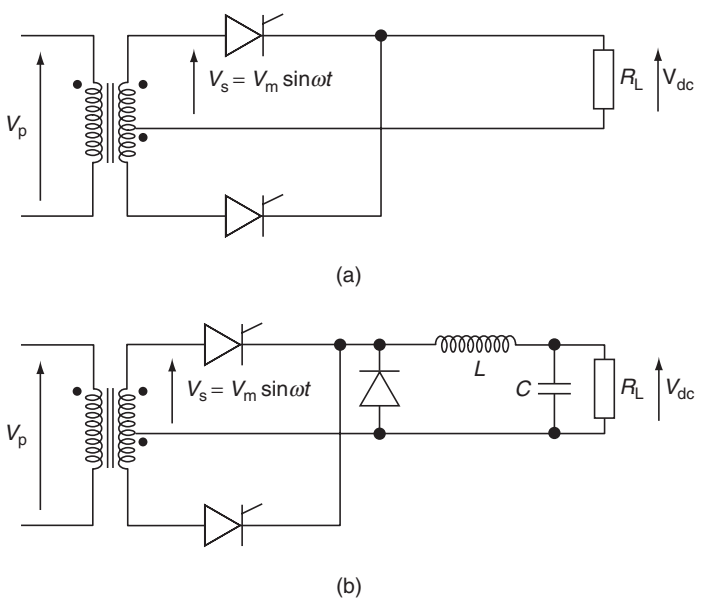
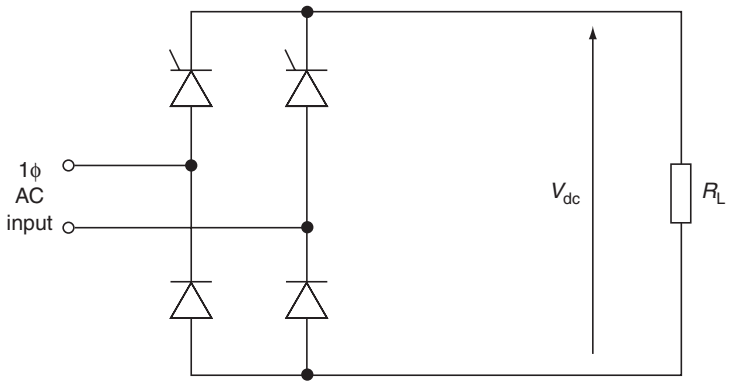
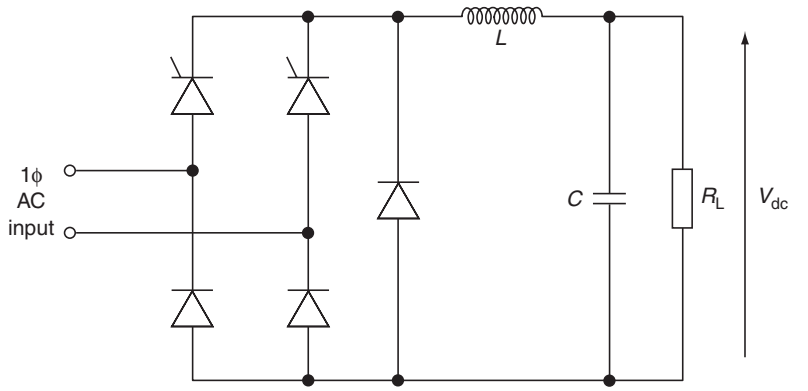


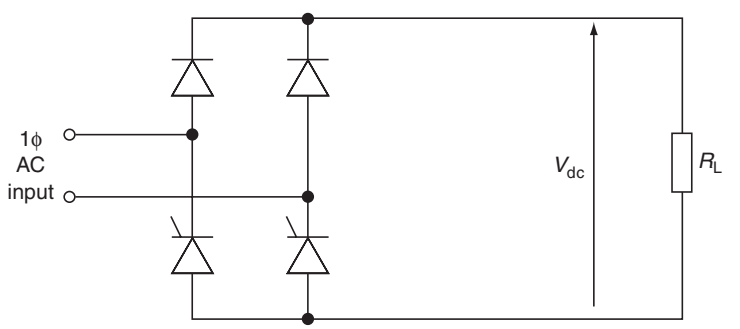
Figure 3.23 | Single-phase, phase-control circuits.



(c)

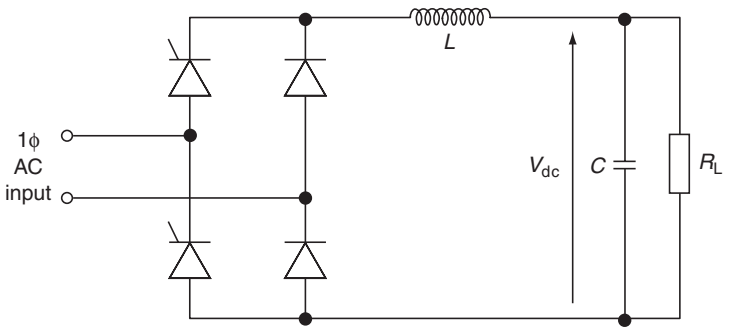


(d)

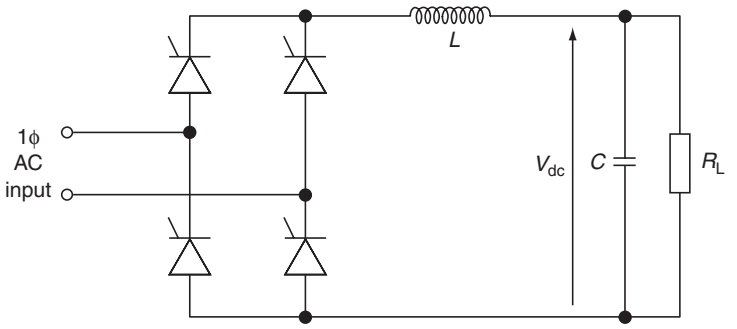


(e)

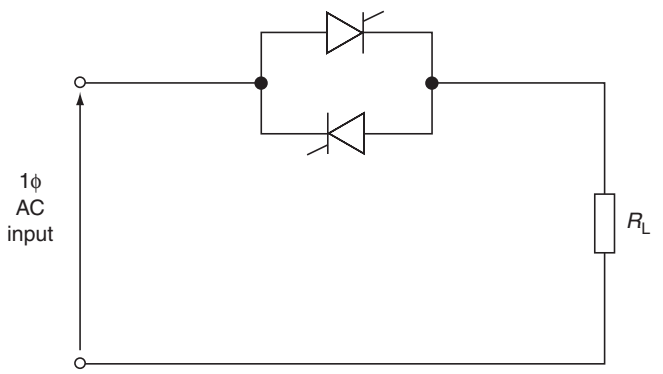
Figure 3.23 | *Continued*.



(f)



(g)



(h)

Figure 3.23 | *Continued*.

$$V_{DC} = \frac{1}{\pi}\int_{\alpha}^{\pi} V_m \sin\theta \, d\theta = \frac{V_m}{\pi}(1+\cos\alpha) \tag{3.71}$$

The outputs of circuits shown in Figures 3.23(a), (c), (e) have an average value as given in Eq. (3.71). The outputs of the rectifier for circuits in Figures 3.23(b), (d), (f) also deliver an average value given above. Here the outputs are filtered using an LC filter to remove the AC component in the rectified output. Circuits of Figures 3.23(b), (d), (f) have a freewheeling diode across the rectifier. Because of this, the rectifier output will contain only positive portions, as the freewheeling diode will clamp the output to zero when it conducts.

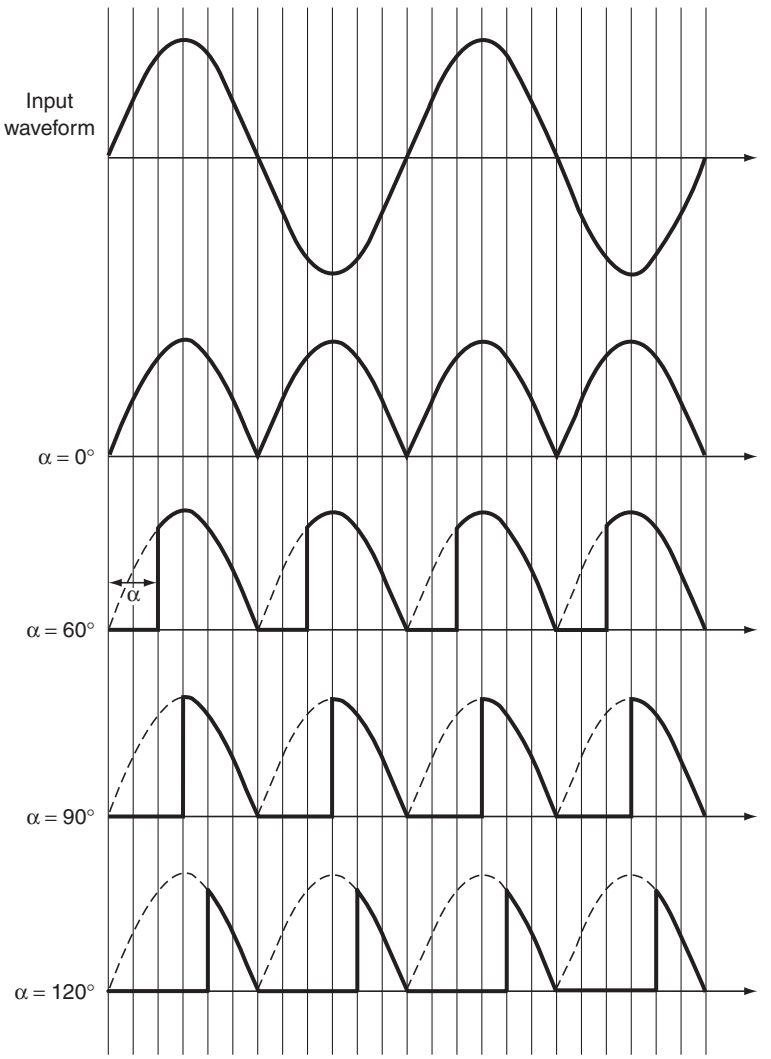


Figure 3.24 Voltage waveforms after rectification for various firing angles for circuits given in Figures 3.23(a)–(f).

Inductive Load

The circuit shown in Figure 3.23(g) has no freewheeling diode. Therefore, SCR commutation will take place only when the inductor current goes below the holding current value. The waveforms for the phase-control circuit of Figure 3.23(g) for various trigger angles are shown in Figure 3.25.

The output voltage with an inductive load is given as

$$V_{DC} = \frac{1}{\pi}\int_{\alpha}^{\pi+\alpha} V_m \sin\theta \, d\theta = \frac{2V_m}{\pi}\cos\alpha \tag{3.72}$$

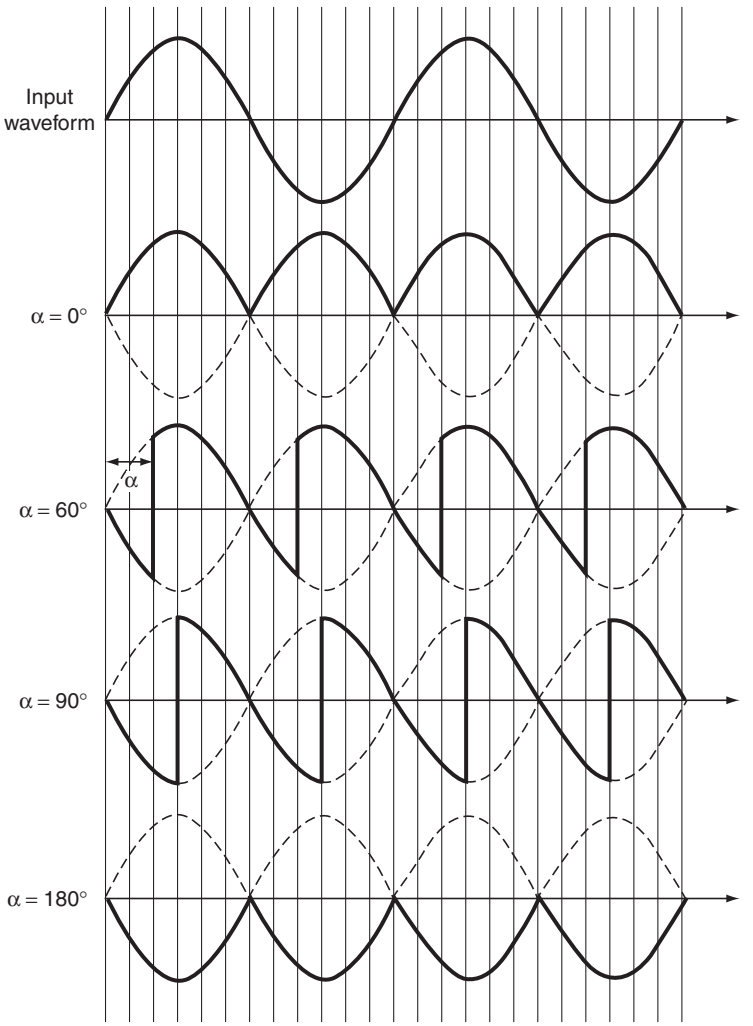


Figure 3.25 Voltage waveforms after rectification for various firing angles for circuit given in Figure 3.23(g).

Figure 3.23(h) provides a topology wherein two SCRs are connected back to back. This topology is a useful circuit for voltage control of resistive loads like heating applications and light dimmer applications. In low-power applications, the two SCRs may be replaced by a single TRIAC to control the voltage to the load as in conventional light dimmers. For high-power applications including inductive loads, the parallel back-to-back SCRs offer additional current capability and overcome the problem of low-commutating dv/dt rating associated with TRIACs. If the resistive load is replaced by a transformer whose secondary voltage is rectified and filtered, a phase-controlled power supply results and regulation is achieved by controlling the firing angle via feedback circuitry from the DC output to the TRIAC or SCR gates.

Three-Phase-Controlled Rectifier Circuits

Various three-phase-controlled power circuits are shown in Figure 3.26. The half-wave circuits of Figures 3.26(a) and (b) are seldom used due to poor utilization of transformer VA ratings. The bridge circuits of Figures 3.26(c)–(f) are more common.

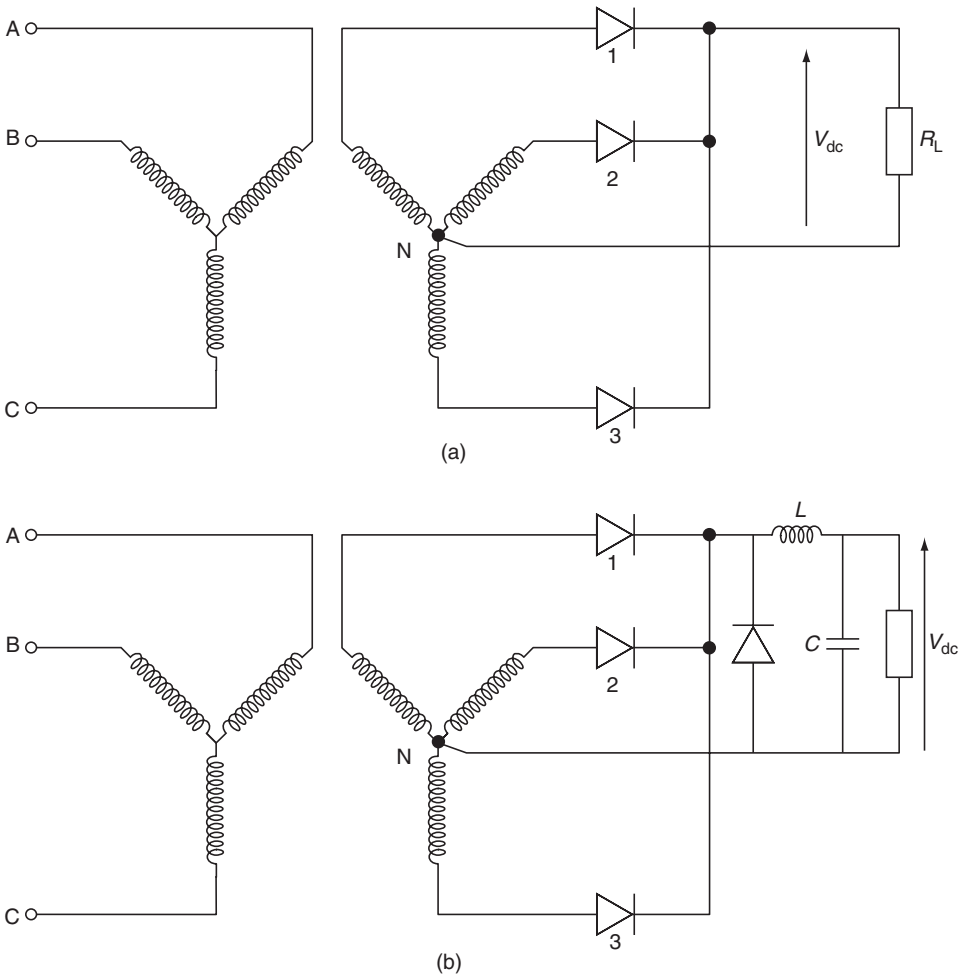
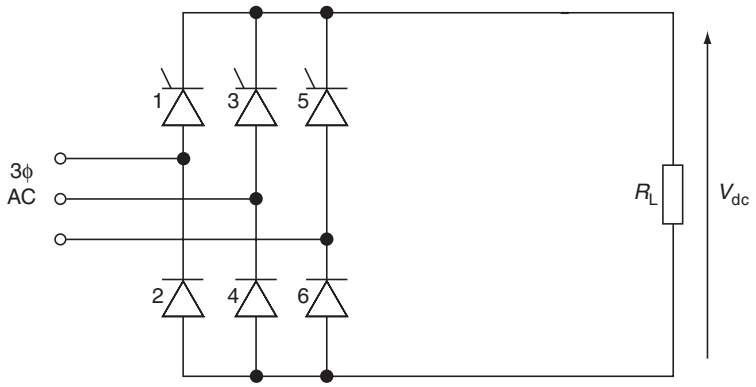
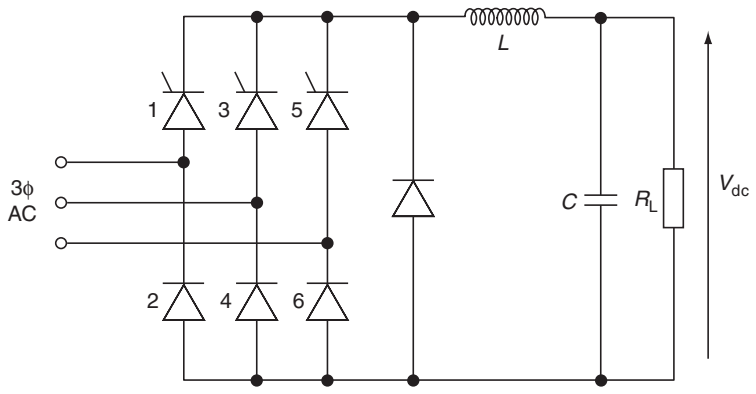


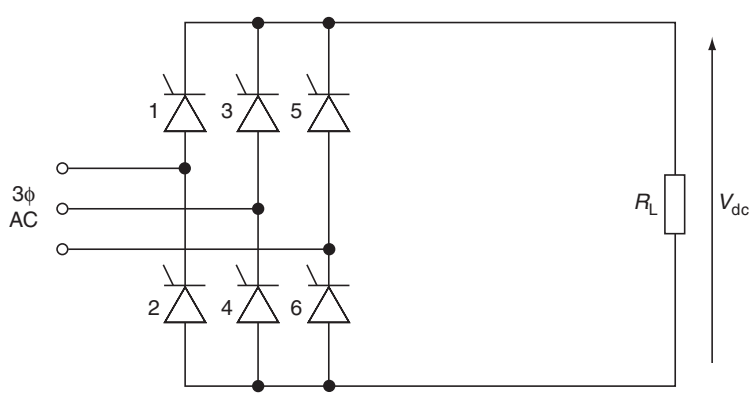
Figure 3.26 | Three-phase, phase-controlled circuits.



(c)



(d)



(e)

Figure 3.26 | *Continued*.

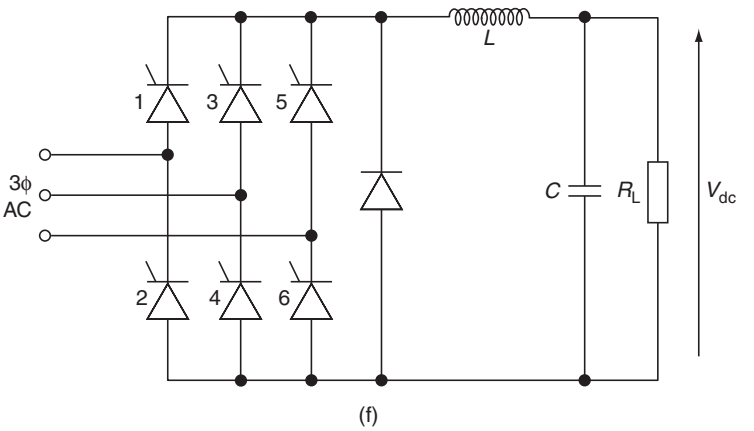


Figure 3.26 | *Continued*.

Half-Wave Control

The half-wave circuits of Figures 3.26(a) and (b) give output voltage waveforms as shown in Figure 3.27. Referring to Figure 3.27, the average or the DC component of the half-wave control circuit output voltage is given as

$$V_{DC} = 3\frac{1}{2\pi}\int_{(\pi/6)+\alpha}^{\pi} V_m \sin\theta \, d\theta = \frac{3V_m}{2\pi}\cos(\alpha + 30^\circ) \tag{3.73}$$

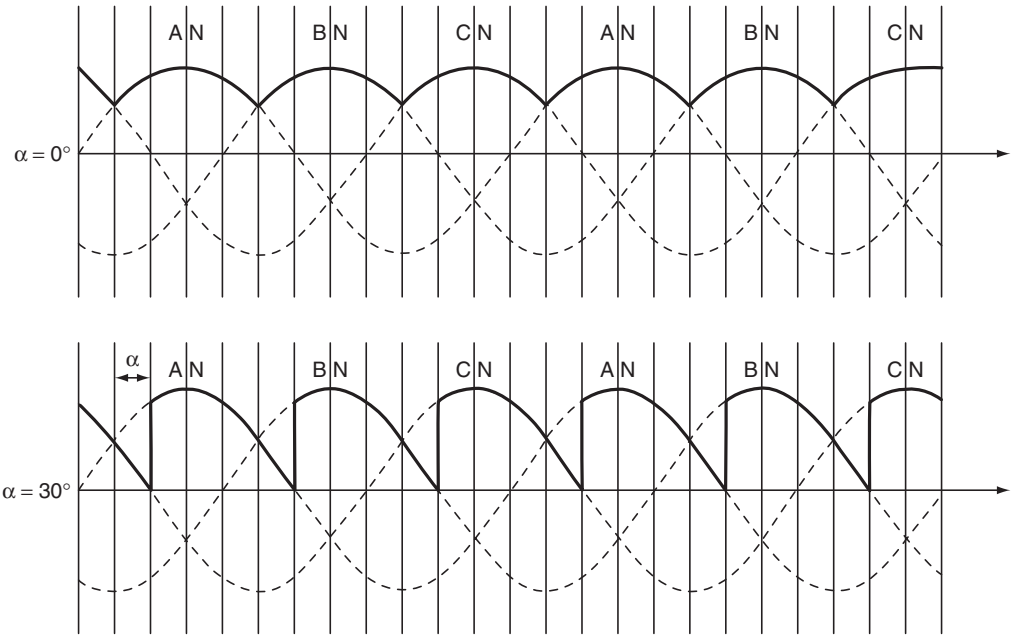


Figure 3.27 | Voltage waveforms after half-control rectification for various firing angles for circuits given in Figure 3.26(a) and (b).

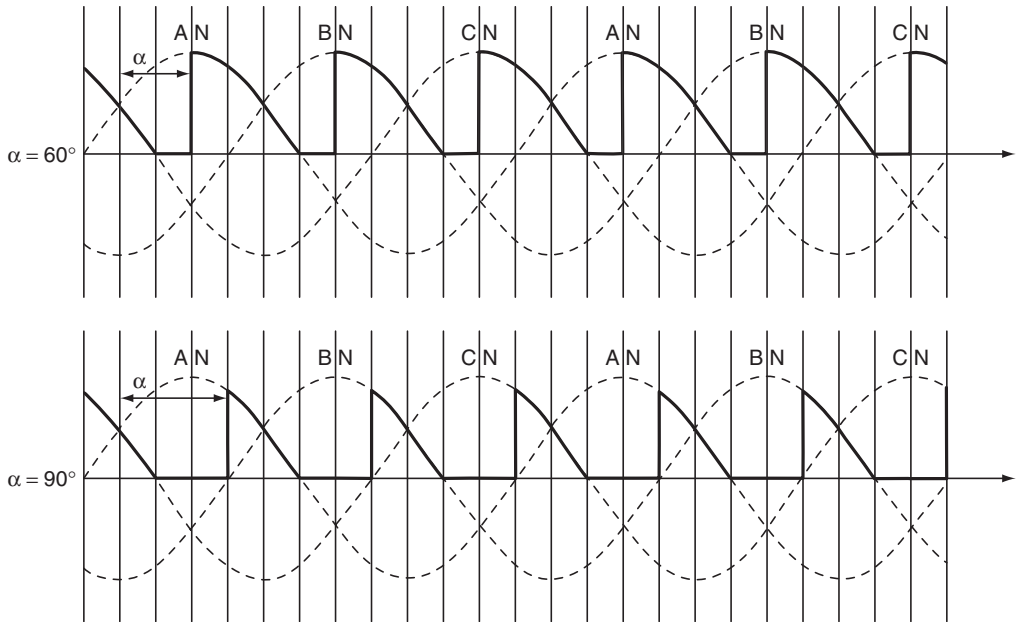


Figure 3.27 | *Continued*.

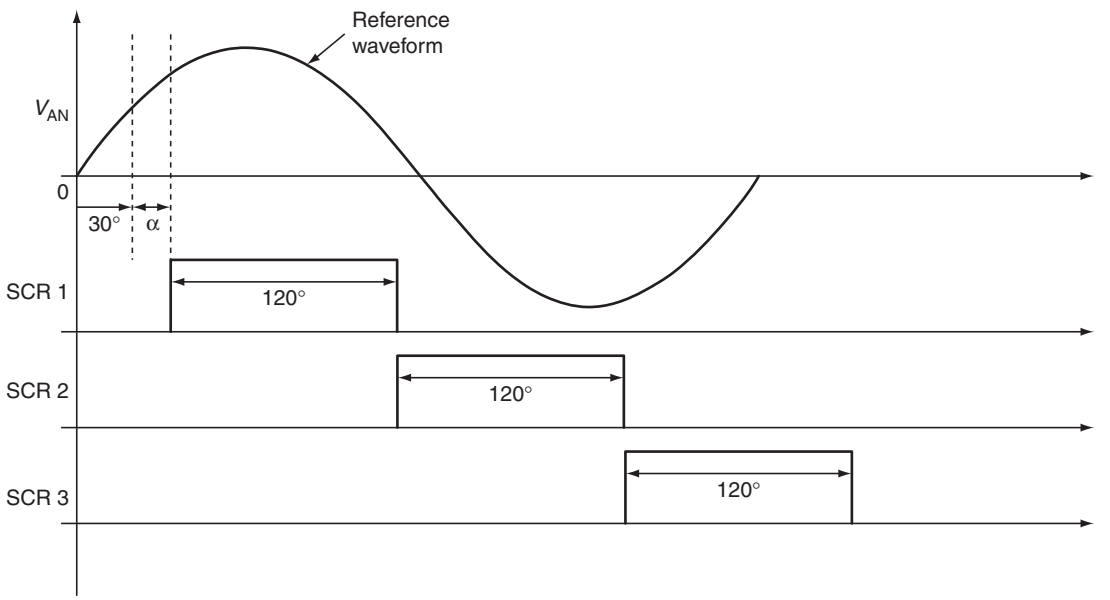


Figure 3.28 | Sequence of SCR firing.

where V_m is the line-to-neutral peak voltage. The sequence of SCR firing for the circuits of Figures 3.26(a) and (b) is shown in Figure 3.28.

Full-Wave, Half-Control

The circuits in Figures 3.26(c) and (d) are full-wave bridges with half-control (only three SCRs are required) and are suitable for resistive load where no regeneration is needed. The SCR cathodes are common which allows a single non-isolated firing stage for triggering the SCRs. However, the output ripple frequency is reduced from six times the lines frequency to three times the line frequency at delays angles $\alpha > 0$ (refer Figure 3.29). The rectifier output voltage waveforms for full-wave half-control circuits are shown in Figure 3.29 for various firing angles α. The average value of the output voltage for the full-wave half-control circuit is given as

$$V_{DC} = \frac{3V_m}{2\pi}(1+\cos\alpha) \tag{3.74}$$

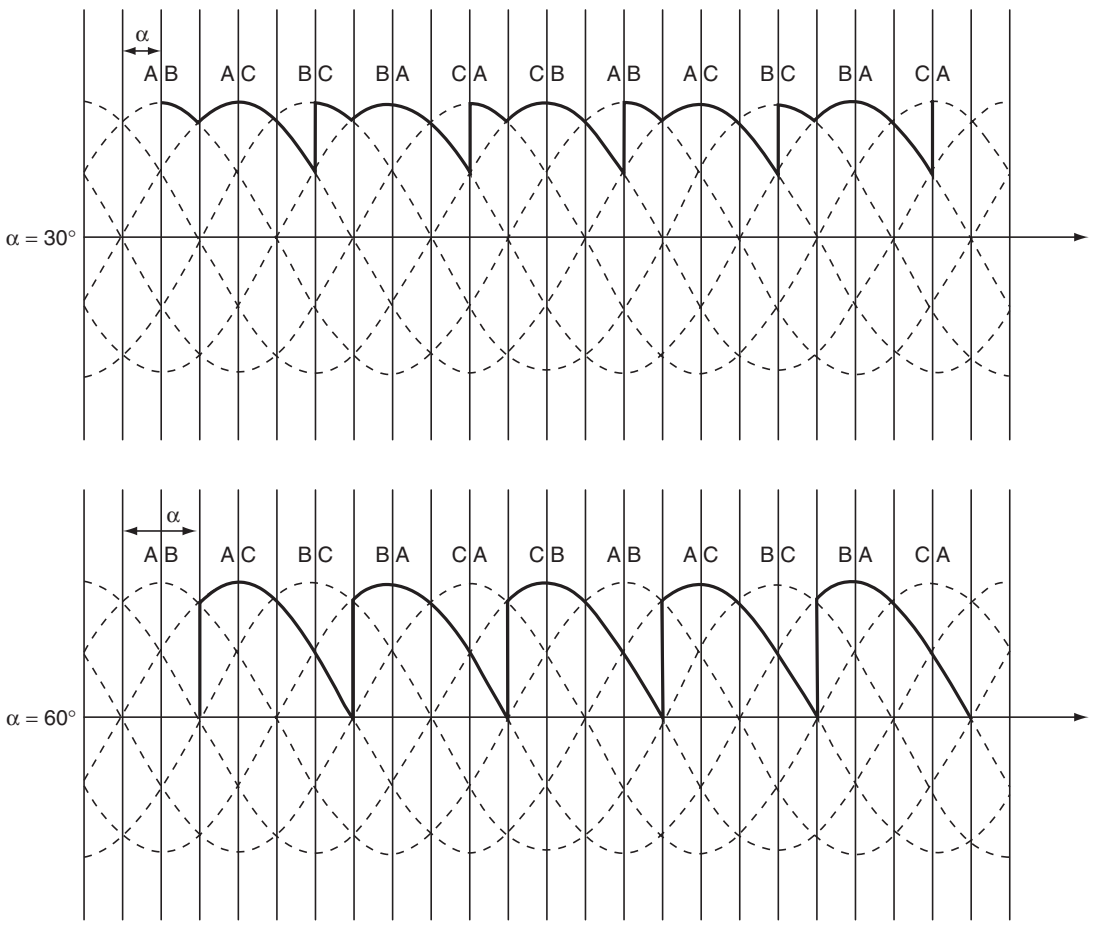


Figure 3.29 Voltage waveforms after rectification for full-wave half-control for various firing angles of circuits given in Figures 3.26(c) and (d).

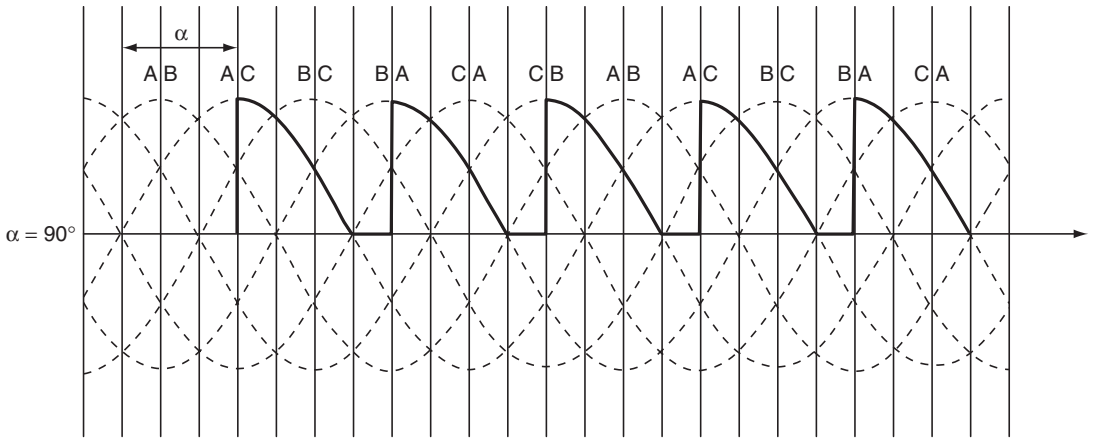


Figure 3.29 | *Continued*.

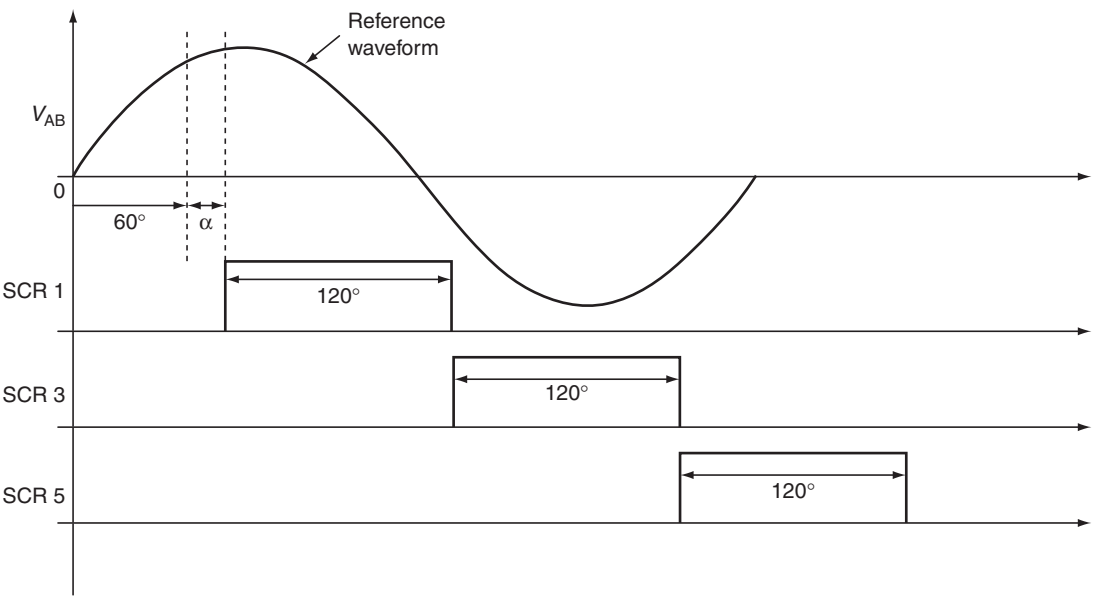


Figure 3.30 | Firing sequence for the SCRs.

where V_m is the line-to-line peak voltage. The firing sequence for the SCRs is as shown in Figure 3.30. The diodes D_2, D_4 and D_6 will conduct depending on the most negative line (A or B or C line) at any given instant of time.

Full-Wave, Full-Control

Full-wave bridges with full control (six SCRs) are shown in Figures 3.26(e) and (f). The firing sequence is more complicated than that of half-control topologies since the SCRs require two gate signals each cycle.

The rectifier output voltage waveforms for various firing angles, α, are shown in Figure 3.31. The average value of the output voltage for the full-wave full-control resistive load is given as

$$V_{DC} = \frac{3V_m}{\pi}\cos\alpha \tag{3.75}$$

where V_m is the line-to-line peak voltage.

The firing sequence for the full-wave, full-control circuits of Figures 3.26(e) and (f) is given in Figure 3.32.

Overlap and Overlap Angle

In the discussions till now, instantaneous commutations among SCRs were assumed. However, in practice this will not be the case. In series with each SCR there will be an inductance. This inductance exists because

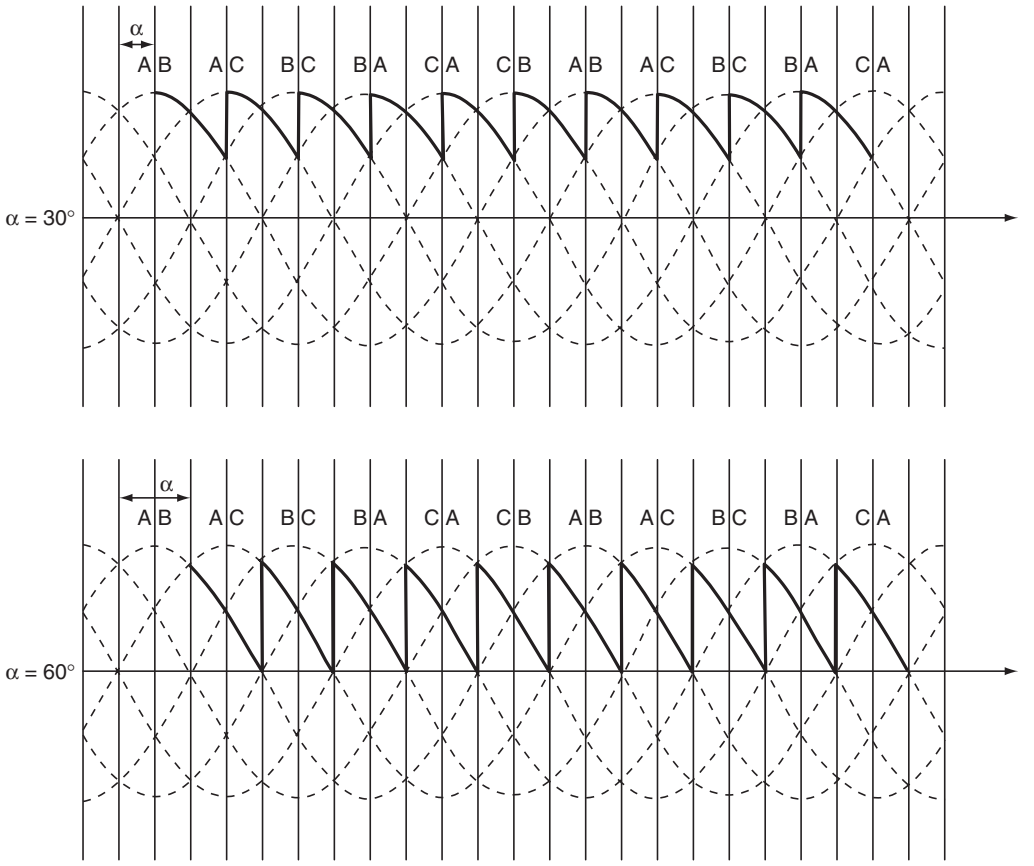


Figure 3.31 Voltage waveforms after rectification for full-wave full-control for various firing angles of circuits given in Figures 3.26(e) and (f).

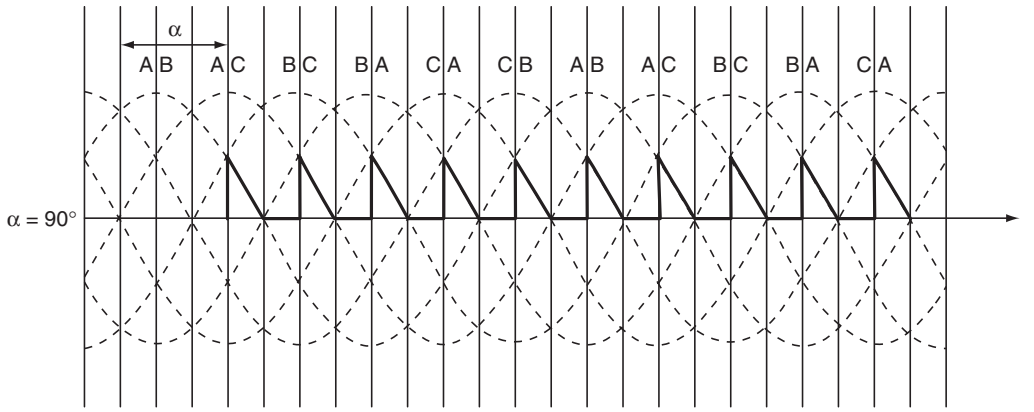


Figure 3.31 | *Continued*.

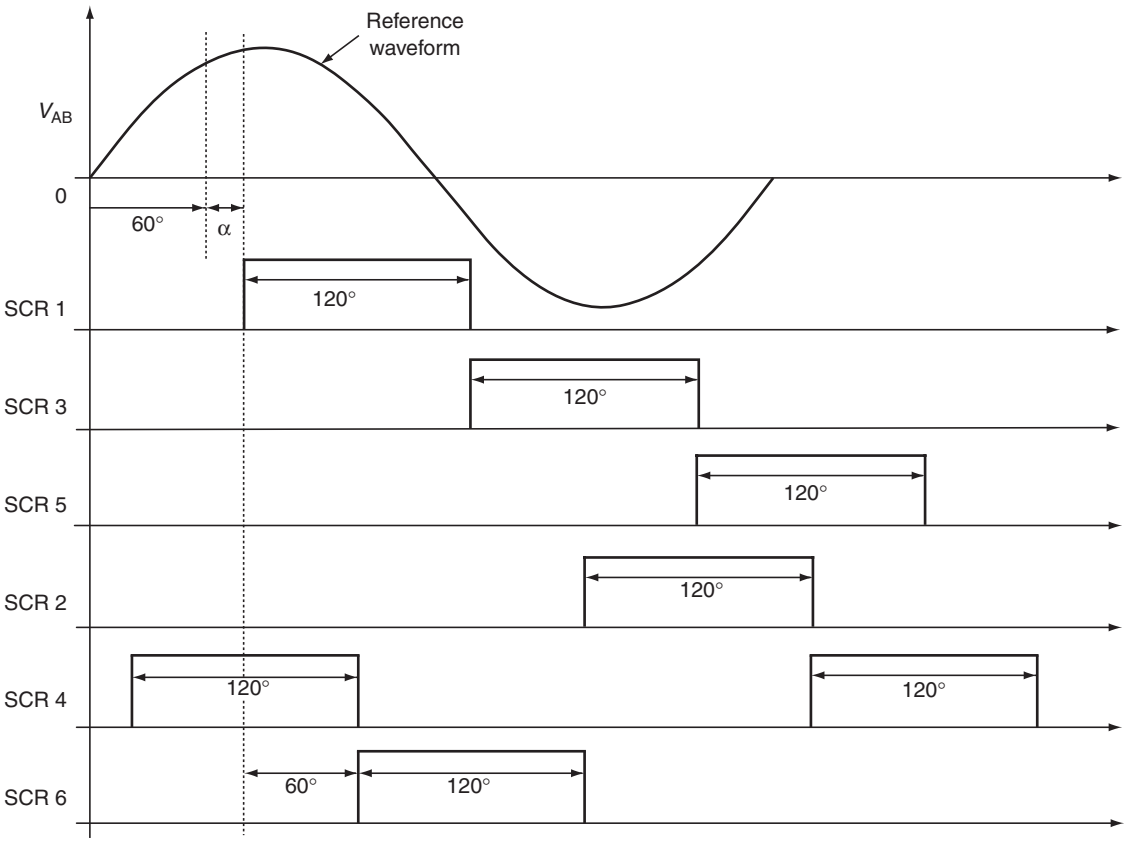


Figure 3.32 | Firing sequence for the full-wave, full-control circuits of Figures 3.26(e) and (f).

of source, line, load, etc. Owing to the presence of this series inductance, during commutation, the SCR that is turning OFF will turn-OFF only after the current through its inductance reduces below the holding current value. The current in the SCR that is turning ON will rise slowly depending on the value of the inductance. During a small interval of the time both the commutating SCRs will be ON. This is called conduction overlap. The period in angle for which the conduction overlap occurs is called the overlap angle. During overlap the output voltage reduces by the inductance voltage drop equal to $L(\mathrm{d}i/\mathrm{d}t)$ and this increases with load current. It should be borne in mind that the presence of these series parasitic inductances that cause the conduction overlap will result in loss of output regulation.

CONCLUDING REMARKS

The rectifiers and especially the capacitor-filter rectifiers are rather ubiquitious as far as power electronic systems are concerned. Most power electronic systems draw the power from the AC grid. They are first converted to DC by the rectification mechanisms discussed in this chapter to form the DC bus or the DC link for the system. In this respect, the AC–DC rectifiers play a pivotal role in most power electronic applications.

Though the capacitor-filter rectifier is volumetrically the most efficient and also least expensive, it suffers from low *pf* issues. This implies that the grid power from which the power is drawn has to be rated many more times than the actual active load power. This leads to significant line loss. Further as the peak currents flow in a small duration near the voltage peaks, the line drops due to the current flow distorts the voltage waveforms causing a flattening of the sinusoidal voltage waveshape near the peaks. This leads to higher total harmonic distortion and conducted electromagnetic interference (EMI) issues. Therefore due to regulations, more and more power electronic systems are incorporating unity *pf* correction circuits into the frontend AC–DC rectifier circuits. The more popular frontend boost converter for *pf* correction and the frontend converters for inverter applications are discussed in Chapter 10.

Apart from the above problems, the issue of the turn-ON surge currents that is prevalent in capacitor-filter rectifier circuits is a serious drawback. Extra circuits have to be used to circumvent the problem of startup surge currents. Notwithstanding the drawbacks mentioned above, the capacitor-filter rectifier is still the most simple and popular AC–DC rectifier circuit that is in use till date. The rectifier LC filter is not as popular due to the bulky and costly inductor. In higher power circuits, controlled rectification followed by capacitor-filter circuits are used. Here the trigger angle or the firing angle gives a measure of control on the startup surge currents.

LABORATORY EXERCISES

1. Consider the capacitor-filter rectifier circuit as shown in Figure 3.33. The input is a single-phase source that is derived from a 230 V mains. The input may be given from an autotransformer to obtain different input voltage amplitudes.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Simulation in SciLAB
- **c.** Hardware bread-boarding

Tasks for study:

- **(a)** Rig up the circuit as shown in Figure 3.33 and plot V_i and i versus time.
- **(b)** Why is there ringing on the current waveform?
- **(c)** What is the series impedance of the circuit?

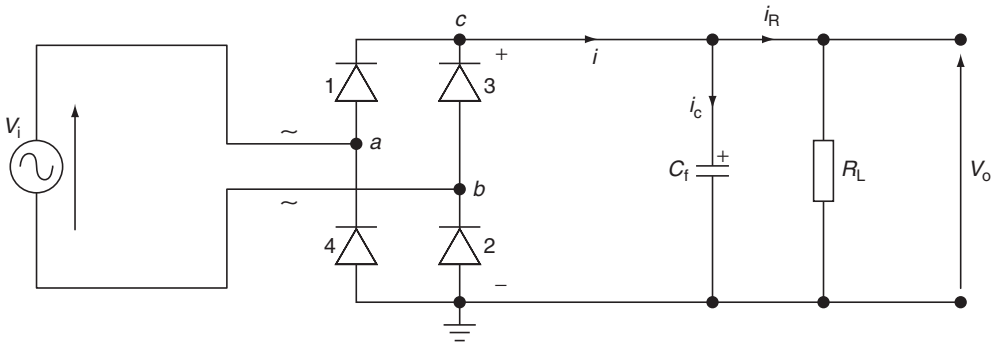


Figure 3.33 | Single-phase capacitor-filter rectifier circuit.

(d) Plot V_i and V_o versus time. What is the ripple? How does ripple depend on load, capacitor C_f and frequency of input waveform?

(e) Measure the current and voltage waveform across the rectifier diode.

(f) Estimate the average and rms currents through the rectifier diode. Calculate the diode power dissipation.

(g) What should be the peak current rating of the diode?

(h) Change the initial charge voltage on the capacitor C_f. What is the effect on the input surge current?

(i) Change the phase angle of the input V_i at start up. Observe the effect on current i. What happens and why?

(j) Under what conditions do you get maximum peak current to flow through the diodes?

(k) Measure the input voltage and current and estimate the power factor.

(l) What is the effect of change in C_f value on the input current and power factor?

2. Consider the three-phase capacitor-filter rectifier circuit as shown in Figure 3.34. The input is a three-phase source that is derived from a 400 V mains. The input may be given from a three-phase autotransformer to obtain different input voltage amplitudes. The circuit shown consists of a basic single-phase full-bridge rectifier. When switches "S_1" are put in position 1, then one more diode arm gets connected. This transforms the circuit into a three-phase full-bridge

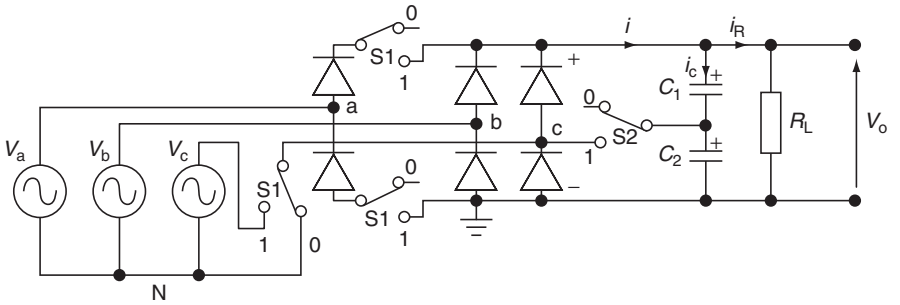


Figure 3.34 | Three-phase capacitor-filter rectifier circuit.

rectifier circuit. Note that when "S_1" are in position 0, V_a and V_c are disconnected by S_1 thereby applying only a single-phase voltage across the b-c diode bridge. During single-phase operation of the circuit, switch "S_2" gives a doubler effect.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Simulation in SciLAB
- **c.** Hardware bread-boarding

Tasks for study:

- **(a)** Keep S_1 in position 0 and S_2 in position 0. Observe V_{bc}, i, i_R, i_C and V_o.
- **(b)** Keep S_1 in position 0 and S_2 in position 1. Observe V_{bc}, i, i_R, i_C and V_o. What is the effect of switch S_2? What is its application?
- **(c)** Keep S_1 in position 1 and S_2 in position 0. Observe V_a, V_b, V_c, i, i_R, i_C and V_o.
- **(d)** What is the ripple frequency of the output when S_1 is in position 1 and in position 0?
- **(e)** When S_1 is in position 1, what is the worst case ripple?
- **(f)** What should be the peak current rating of the diode?
- **(g)** Change the initial charge voltage on the capacitors. What is the effect on the turn-ON surge current?
- **(h)** Observe the diode currents when S_1 is in position 1 and in position 0.
- **(i)** Estimate the average and rms currents of the diode when S_1 is in position 1 and in position 0.
- **(j)** What is the effect of the S_2 switch on the diode currents for a given load?
- **(k)** What is the effect of unequal C_1 and C_2 on the output voltage?
- **(l)** How is charge equalization done for C_1 and C_2? [Hint: Connect resistors across each capacitor and observe.]

3. Consider the single-phase LC-filter rectifier circuit as shown in Figure 3.35. The input is a single-phase source that is derived from a 230 V mains. The input may be given from an auto-transformer to obtain different input voltage amplitudes. L_s and R_s are the equivalent circuit series inductance and resistance, respectively, that represent the non-idealities; L_f is the filter inductance of the LC filter.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Simulation in SciLAB
- **c.** Hardware bread-boarding

Tasks for study:

- **(a)** Rig up the circuit as shown above and plot V_i and i versus time.
- **(b)** Plot the rectified V_i and V_o versus time.
- **(c)** Vary the load resistor R_L and plot V_o versus load I_o.
- **(d)** What is the minimum load current for which the inductor current is continuous?
- **(e)** Incorporate a bleeder resistor across the capacitor and plot the efficiency versus load with and without the bleeder resistor.

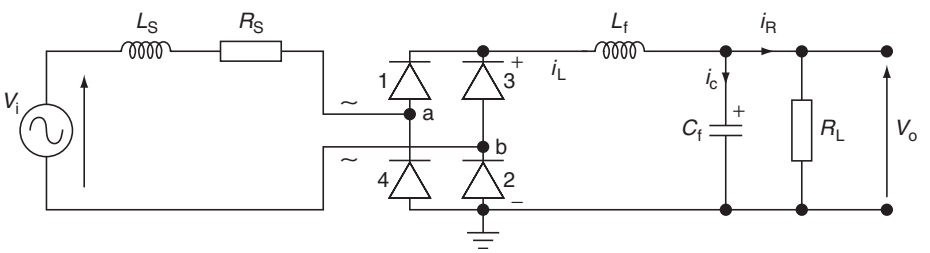


Figure 3.35 | Single-phase LC–rectifier circuit.

(f) What is the ripple? How does a ripple depend on load, inductor, capacitor and frequency of input waveform?
(g) Measure the current and voltage waveforms across the rectifier diode.
(h) Estimate the average and rms currents through the rectifier diode. Calculate the diode power dissipation.
(i) What should be the peak current rating of the diode?
(j) Compare the turn-ON surge current with the capacitor-filter rectifier for a given load.
(k) Measure the input voltage and current and estimate the power factor.
(l) What is the effect of change in L_f and C_f value on the input current and power factor?

FILL IN THE BLANKS

1. Single-phase rectifier circuit topologies are designed for applications wherein the input source is the _____.

2. Three-phase rectifier circuit topologies are designed for applications wherein the input source is a _____.

3. For inductive loads, an additional diode called _____ is needed to provide a path for the inductor current when the rectifier diode is OFF.

4. Use of a _____ is compulsory for the full-wave center-tapped rectifier.

5. The PIV for the diodes of a full-wave center-tapped rectifier is _____ the PIV for the diodes of the half-wave or full-bridge rectifiers for a given load voltage requirement.

6. In a full-wave center-tapped rectifier, the current through the secondary of the transformer is a _____ wave rectified waveform.

7. In a full-wave center-tapped rectifier, the current through the rectifier diodes of the transformer is a _____ wave rectified waveform.

8. In the full-wave bridge rectifier there will be _____ diode drop in the bridge rectifier topology as compared to the half-wave or full-wave center-tapped rectifier topologies.

9. In a full-wave bridge rectifier with transformer, the current through the secondary of the transformer is a _____ waveform.

10. In a full-wave bridge rectifier, the current through the rectifier diodes is a _____ wave rectified waveform.

11. In the case of the bridge rectifier circuit, there is no need for a separate _____ diode.

12. The load ripple frequency for a half-wave rectifier system is _____ the input source frequency.

13. The load ripple frequency for a full-wave rectifier system is _____ the input source frequency.

14. In a three-phase half-wave rectifier circuit, the secondary windings of the transformer carry the _____ currents as that flowing through the diodes.

15. The load ripple frequency for a three-phase half-wave rectifier system is _____ the input source frequency.

16. The load ripple frequency for a three-phase full-wave rectifier system is _____ the input source frequency.

17. Capacitor input filters are the most _____ efficient means of filtering rectified sine waves and storing energy.

18. The output voltage ripple for a single-phase capacitor-filter rectifier is _____ the input frequency.

19. The output ripple frequency for three-phase half-wave capacitor-filter rectifier is _____.

20. The output ripple frequency for three-phase full-wave capacitor-filter rectifier is _____.

21. In a three-phase capacitor-filter rectifier circuit that is supplied from three-phase mains, the

peak capacitor voltage is _____ time the input mains phase voltage.

22. The turn-ON surge current in the capacitor-filter rectifier is limited by the line resistance, _____ of the capacitor, rectifier diodes _____ and the input line and/or the leakage _____ of the transformer if present.

23. Low power factor implies _____ line losses.

24. Unity power factor occurs when the load is purely _____.

25. In a capacitor-filter rectifier circuit, lower the power factor, _____ will be the _____ current drawn from the input source.

26. The rectifier–LC filter does not have the problem of turn-ON _____ due to the presence of the _____.

27. The major drawback of the LC-filter rectifier is the _____ that has to be designed for the supply frequency.

28. For a good LC-filter circuit, _____ and $X_c << X_L$.

29. In LC–rectifier filters, when the load is light, a _____ value of L is needed if the ripple is to be small.

30. Power supplies utilize TRIACs or SCRs as the _____ device to provide a desired dc output from single-phase or three-phase inputs.

31. In single-phase half-controlled SCR bridge rectifier configurations, the rectifier output will contain only _____ portions.

32. In single-phase full-controlled SCR bridge rectifier configuration, the SCR commutation will take place only when the inductor current goes below the _____ value.

33. The reference waveform for sequencing the firing of the SCRs for the three-phase half-wave half-controlled rectifier is the _____ waveform.

34. The reference waveform for sequencing the firing of the SCRs for the three-phase full-wave half-controlled rectifier is the line-to-line _____ waveform.

35. Three-phase full-wave full-controlled rectifier consists of _____ SCRs in bridge configuration.

DESCRIPTIVE QUESTIONS

1. Where are rectifiers used?

2. How are rectifiers classified?

3. What are the three basic rectifier configurations?

4. For a half-wave rectifier, draw the load voltage and current waveforms for an inductive load and derive the load power equation.

5. For a full-wave center-tapped rectifier, draw the load voltage and current waveforms for an inductive load and derive the load power equation.

6. For a full-wave bridge rectifier, draw the load voltage and current waveforms for an inductive load and derive the load power equation.

7. Explain the operation of the three types of single-phase rectifier circuits with illustrative waveforms.

8. A delta-star transformer is used for a three-phase full-wave rectifier system. In the secondary of the three-phase transformer, if the neutral point is available, draw and explain the schematic to generate positive and negative voltages with respect to the neutral point using a single three-phase bridge-rectifier topology.

9. Discuss the operation of the capacitor input filter rectifier.

10. What are the factors that affect the output ripple of the capacitor-filter rectifier?

11. What is hold time? How does one design the value of capacitor for a specified hold time?

12. Discuss the turn-ON surge current problem in capacitor-filter rectifier circuits. What are the parameters affecting the amplitude of the turn-ON surge current? How is this problem solved in practice?

13. What are the methods for improving the power factor of AC–DC rectifiers?

14. What are the benefits of the LC–rectifier filter?

15. What are the disadvantages of the LC–rectifier filter?

16. Discuss the design considerations for the inductor of an LC–rectifier filter.

17. Explain the terms minimum load current and bleeder load resistance with respect to a rectifier–LC filter.

18. What is the ripple content in the output voltage of an LC–rectifier filter?

19. Discuss the operation of half- and full-control of single-phase controlled rectifiers.

20. Distinguish between half-wave half-control, full-wave half-control and full-wave full-control in three-phase-controlled rectification.

PROBLEMS

1. A half-wave rectifier is supplying 50 W to a resistive load by drawing power directly from a 230 V rms mains grid. What are the average load voltage and current? What are the rms load voltage and current?

2. A full-wave center-tapped rectifier is supplying 50 W to a resistive load by drawing power from a center-tapped transformer that is connected to the 230 V rms mains grid. The turns ratio of the center-tapped transformer is unity. What are the average load voltage and current? What are the rms load voltage and current?

3. For Problem 2, calculate the rms value of the primary and secondary winding currents.

4. A full-wave bridge rectifier is supplying 50 W to a resistive load by drawing power directly from the 230 V rms mains grid. What are the average load voltage and current? What are the rms load voltage and current?

5. For Problem 4, calculate the rms value of the primary and secondary winding currents.

6. A 1000 W resistive load is supplied by a star–star transformer-isolated three-phase half-wave rectifier. The rectifier is drawing power directly from the $400V_{\text{rms}}$ three-phase mains grid. The secondary to primary phase winding turns ratio is 0.5. What are the average load voltage and current? What are the rms load voltage and current? What are the average and rms currents of the rectifier diodes?

7. A 1000 W resistive load is supplied by a star–star transformer-isolated three-phase full-wave rectifier. The rectifier is drawing power directly from the $400V_{\text{rms}}$ three-phase mains grid. The secondary to primary phase winding turns ratio is 0.5. What are the average load voltage and current? What are the rms load voltage and current? What are the average and rms currents of the rectifier diodes?

8. In a capacitor-filter rectifier that is supplied from 230 V, 50 Hz mains, a 100 μF output capacitor discharges to 250 V every half-cycle. What is the energy given to the load by the capacitor every input voltage cycle?

9. Calculate the capacitor value of a capacitor-filter rectifier that is supplied from 230 V, 50 Hz mains wherein the peak-to-peak ripple is 50 V and the load is 500 W.

10. For Problem 9, calculate the load current and the peak, rms and average values of the current

through the rectifier diodes. Calculate also the rms current through the capacitor.

11. For Problem 9, estimate the power factor of the capacitor-filter rectifier circuit.

12. For an LC–rectifier filter that is supplied from the 230 V, 50 Hz mains, the load power is 500 W, what is the output voltage if the current through the inductor is continuous?

13. For an LC–rectifier filter that is supplied from the 230 V, 50 Hz mains, if the maximum value of the load resistance is 50 Ω, estimate the minimum value of the filter inductor required.

14. For an LC–rectifier filter that is supplied from the three-phase 400 V mains, if the maximum value of the load resistance is 50 Ω, estimate the minimum value of the filter inductor required.

15. For Problem 13, if the rms value of the output ripple is not to exceed 50 V, calculate the capacitor value of the LC filter.

16. A single-phase half-controlled SCR bridge rectifier is supplying a resistive load of 1000 W from a single-phase 230 V mains. What is the average value of the load voltage if the SCRs are triggered at firing angles of 45°, 90°, 120° and 180°?

17. A single-phase full-controlled SCR bridge rectifier is supplying an RL load from a single-phase 230 V mains. What is the average value of the load voltage if the SCRs are triggered at firing angles of 45°, 90°, 120° and 180°?

18. For a three-phase half-wave half-controlled rectifier which is supplying a load from a three-phase 400 V rms line-to-line mains grid, find the trigger angle at which the average load voltage is 75V.

19. For a three-phase full-wave half-controlled rectifier which is supplying a load from a three-phase 400 V rms line-to-line mains grid, find the trigger angle at which the average load voltage is 400 V.

20. For a three-phase full-wave full-controlled rectifier which is supplying a load from a three-phase 400 V rms line-to-line mains grid, find the trigger angle at which the average load voltage is 400 V.

ANSWERS

Fill in the Blanks

1. single-phase grid
2. three-phase grid
3. freewheeling diode
4. transformer
5. twice
6. half
7. half
8. two
9. full
10. half
11. freewheeling
12. same as
13. twice
14. same
15. thrice
16. four times
17. volumetrically
18. twice
19. 150 Hz
20. 300 Hz
21. $\sqrt{3}$
22. equivalent series resistance; dynamic resistance; reactance
23. higher
24. resistive
25. higher; peak
26. surge currents; inductor
27. inductor
28. $X_c << R_L$
29. large
30. control
31. positive
32. holding current
33. line-to-neutral A phase
34. V_{AB}
35. six

CHAPTER 4

DC–DC Linear Regulators

Learning Objectives

After reading this chapter, you will be able to:

- understand the operating principle of linear regulators.
- design and apply the various types of linear regulators.
- analyze and characterize the linear regulators.

The DC–DC linear regulators convert the input DC voltage to an output DC voltage wherein the output voltage is regulated for variations in input voltage, temperature and output load. The regulation is achieved by operating the power semiconductor devices in the linear region rather than as switches. This, of course, leads to large power dissipation in the power devices and consequent reduction in efficiency of the regulator.

There are, in general, two component sets in the operation of the DC–DC linear regulators, namely, (a) the voltage scaling circuit and (b) the output voltage regulation circuit. However, the operations of these two component sets are interlinked and dependent on each other. Hence the designs of these two component sets are not decoupled from each other.

The output DC voltage value is always lesser than the input unregulated DC voltage value, that is, the scaling circuit will only attenuate the input voltage but will not amplify the input voltage. In other words, the gain in the scaling circuit is less than unity. Even though the linear regulators have low efficiency, the quality of the output voltage with respect to the variations in parameters like input voltage, temperature and load are excellent and by far better than that offered by any switched-mode converter. The output voltage can be designed to be almost ripple-free in the case of the linear regulators. The major disadvantage of the linear regulators is in term of efficiency and as a consequence the very low resulting volumetric power density.

4.1 Generic Linear Regulator

The schematic of a generic linear regulator is shown in Figure 4.1. It consists of two main resistive components: (a) R_s, the series component that performs the function of dropping the voltage across itself to obtain a specified scaling at the output; (b) R_{sh}, the shunt component that generally performs the function of the regulation. Both R_s and R_{sh} are variable resistances.

R_s is an essential component is all linear regulators. However, R_{sh} may not be present in all linear regulator topologies. If the output voltage regulation is performed by varying R_s, such regulators are classified as series regulators and if it is performed by varying R_{sh} then such regulators are classified as shunt regulators.

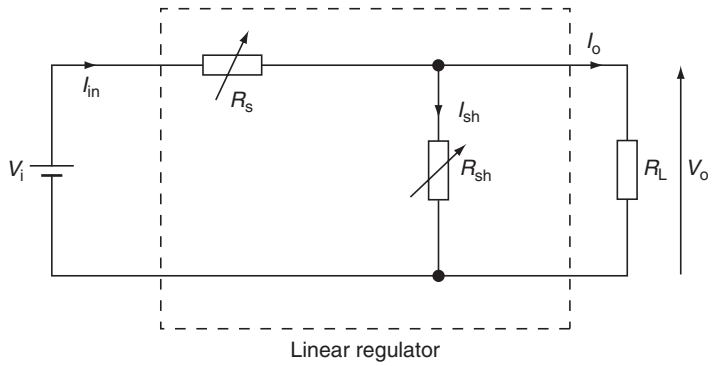


Figure 4.1 | Schematic of a generic linear regulator.

Operating Principle

The operation of the generic linear regulator depends on whether it is the series regulator type or the shunt regulator type. In the series regulator, as the output voltage V_o increases, the series resistance R_s is increased. The drop across R_s increases thereby bringing down the V_o. Likewise as V_o decreases, R_s is decreased. The drop across R_s decreases and brings up V_o. In the shunt regulator, the shunt component is used for regulating V_o. Referring to Figure 4.1, one can observe that

$$I_{in} = I_{sh} + I_o = \frac{V_o}{R_{sh}} + \frac{V_o}{R_L} \tag{4.1}$$

$$V_o = V_i - I_{in} R_s \tag{4.2}$$

Let P_i be the power drawn from the input source. Using Eqs. (4.1) and (4.2), P_i is given as

$$\begin{aligned} P_i &= V_i I_{in} = (V_o + I_{in} R_s) I_{in} \\ &= V_o I_{in} + I_{in}^2 R_s \\ &= V_o I_o \left(\frac{R_{sh} + R_L}{R_{sh}} \right) + I_o^2 \left(\frac{R_{sh} + R_L}{R_{sh}} \right)^2 R_s \\ &= V_o I_o \left(1 + \frac{R_L}{R_{sh}} \right) + I_o^2 \left(1 + \frac{R_L}{R_{sh}} \right)^2 R_s \end{aligned} \tag{4.3}$$

The output power (P_o) delivered to the load is

$$P_o = V_o I_o \tag{4.4}$$

The efficiency of the linear regulator is defined as the ratio of the power delivered to the output load to the power drawn from the input source. This is given as

$$\eta = \frac{P_o}{P_i} = \frac{V_o I_o}{V_o I_o [1 + (R_L / R_{sh})] + I_o^2 [1 + (R_L / R_{sh})]^2 R_s} \tag{4.5}$$

If R_{sh} is very high or tending to infinity, then $I_{in} = I_o$. From Eq. (4.5) it can be observed that

$$\eta = \frac{V_o}{V_i}$$

4.2 Linear Regulator Topologies

Shunt Regulator

Figure 4.2 shows the circuit schematic of a Zener shunt regulator. The shunt component is a Zener diode that is operated in the reverse breakdown region. The series component is a fixed resistor R_s.

Ideal Zener

Figure 4.3 shows the static characteristic of the Zener diode. The Zener diode is the shunt element that is connected across the load. According to the idealized static characteristic of the Zener diode as shown in Figure 4.3(a), one can observe that the voltage across the Zener diode is a constant at V_z. Thus V_o is clamped to V_z irrespective of the changes in the input voltage and the load. Thus V_o is regulated by virtue of the Zener diode being operated in the breakdown region. Referring to Figure 4.2,

$$I_{in} = I_z + I_o \tag{4.6}$$

$$I_{in} = \frac{V_i - V_z}{R_s} \tag{4.7}$$

$$V_o = V_z \tag{4.8}$$

From Eq. (4.8), it is seen that V_o is independent of the load I_o and the input V_i. Thus the load voltage is regulated as V_z is constant. Under no-load condition, $I_o = 0$ and the whole of I_{in} flows into I_z. As the load increases, a part of I_{in} commutates from the Zener to the load, that is, I_z decreases and I_o increases. Likewise, if I_o decreases, I_z will increase to compensate for the decrease in I_o, thus regulating the output. The Zener

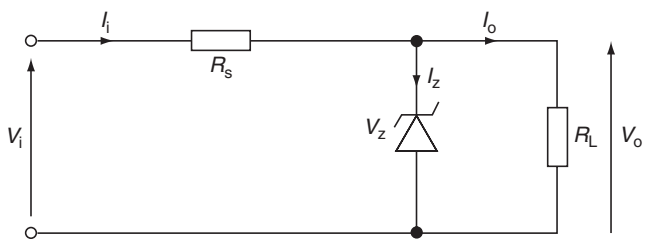


Figure 4.2 | Zener shunt regulator.

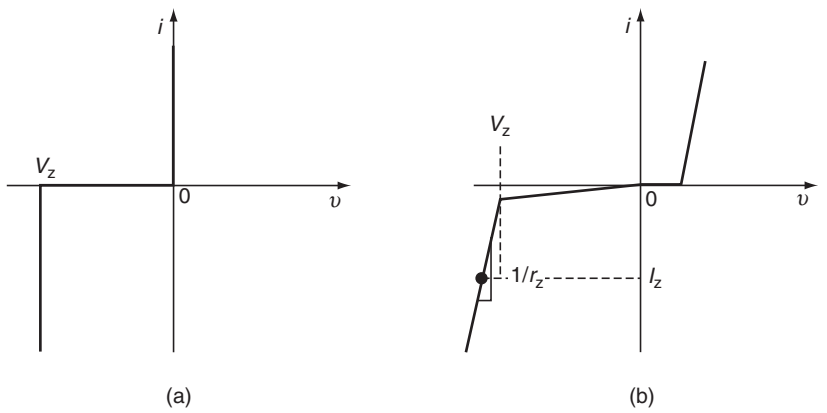


Figure 4.3 (a) Ideal static characteristic of the Zener diode; (b) piece-wise linear static characteristic of the Zener diode.

current I_z is maximum under no-load condition when I_o is zero. Under this condition, the input current I_{in} flows through the Zener. Thus,

$$I_{z\text{-max}} = \frac{V_{i\text{-max}} - V_z}{R_s} \tag{4.9}$$

$I_{z\text{-max}}$ is limited only by the power rating of the Zener diode. The other limiting condition is when the input current completely flows through the load under full-load condition. During this condition,

$$I_o = I_{in} \text{ and } I_z = 0 \tag{4.10}$$

The Zener should be maintained in the breakdown region even under minimum input voltage condition. Therefore, the maximum load current allowable is

$$I_{in} = I_{o\text{-max}} = \frac{V_{i\text{-min}} - V_z}{R_s} = \frac{V_z}{R_{L\text{-min}}} \tag{4.11}$$

Problem 4.1

Consider a shunt regulator circuit wherein the input voltage V_i varies between 9 and 13 V and the Zener diode used is a 5 V, 400 mW device. What is the range of the load resistance?

Solution

The maximum possible Zener current is

$$\frac{P_z}{V_z} = \frac{400 \text{ mW}}{5 \text{ V}} = 80 \text{ mA}$$

The operating input current should be less than 80 mA. Let 60 mA be the maximum Zener operating current $I_{z\text{-max}}$. Then from Eq. (4.9),

$$R_s = \frac{V_{i\text{-max}} - V_z}{I_{z\text{-max}}} = \frac{13-5}{60\text{ mA}} \approx 133\,\Omega$$

The power dissipated in the series resistor R_s is

$$P_{Rs} = \frac{(13-5)^2}{133} \approx 0.48\text{ W}$$

From Eq. (4.11), the minimum value of the load resistance R_L should be calculated such that even at the lowest input voltage and load, the Zener is in reverse breakdown. Thus,

$$R_{L\text{-min}} = \frac{R_s V_z}{V_{i\text{-min}} - V_z} = \frac{133\times 5}{9-5} \approx 166\,\Omega$$

On the other hand, the maximum value of the load resistance can be infinite, that is, open circuit when all the input current flows through the Zener diode. Thus the range of the load resistance is limited to

$$166 \le R_L < \infty$$

Non-Ideal Zener

The Zener diode does not have ideal static characteristics as shown in Figure 4.3(a). The piece-wise linear approximation of the actual static characteristic of the Zener diode shown in Figure 4.3(b) is much closer to reality. The Zener has a non-zero operating resistance r_z in the breakdown region as shown in Figure 4.3(b). This first-level non-ideality is introduced into the shunt regulator circuit as indicated in Figure 4.4.

Owing to the presence of r_z, the output voltage is no longer identical to the Zener breakdown voltage V_z. It is now given as

$$V_o = V_z + I_z r_z \tag{4.12}$$

From Eq. (4.12) it is evident that the output voltage is dependent on I_z. However, I_z varies from almost zero at full-load condition to full load current value during no-load operation. Therefore, the regulation of the output with respect to variations in the load is poor due to the voltage drop across r_z.

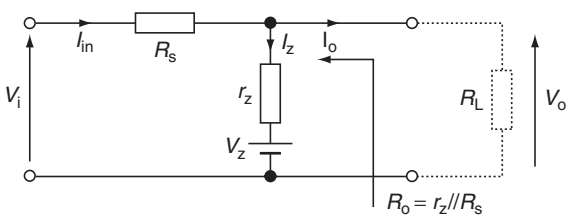


Figure 4.4 The Zener diode is replaced by first-level non-idealities.

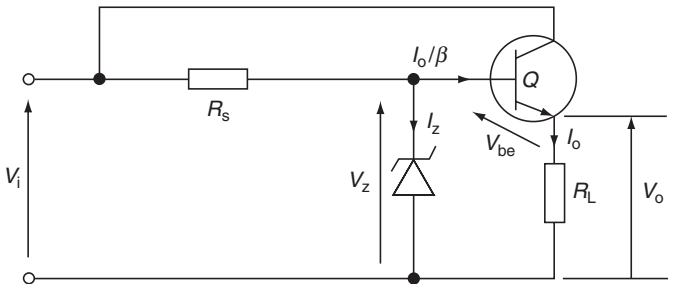


Figure 4.5 | Modified shunt regulator.

Modified Shunt Regulator

The regulator shown in Figure 4.5 tries to reduce the effect of the Zener diode non-ideality r_z on the output voltage regulation.

For the modified shunt regulator of Figure 4.5, the output voltage V_o is the Zener voltage V_z minus the drop across the base–emitter junction of the transistor Q. Thus,

$$V_o = V_z - V_{be} \tag{4.13}$$

In this circuit, the maximum Zener current is limited to I_o/β instead of I_o for the circuit of the Figure 4.4. As the Zener current variation is reduced by β, the drop across the Zener resistance r_z is also very much diminished. Thus, the output voltage including the Zener non-ideality is given as

$$V_o = V_z + I_z r_z - V_{be} \tag{4.14}$$

where maximum value of I_z is I_o/β.

In this topology, even though the output voltage regulation with respect to variations in the load is much better as compared to the circuit of Figure 4.4, the outp ut voltage is dependent on V_{be} of the transistor Q. The V_{be} of the transistor varies with temperature at the rate of about −2 mV/°C rise in temperature. However, if the Zener is chosen to have a positive temperature coefficient then this would compensate for the variations in V_{be} due to temperature, thereby making the output voltage V_o independent of temperature.

Series Regulator

The series regulator is a linear regulator wherein the series component is controlled. To control the series component, a controllable resistor has to be used as the series-pass element. A transistor or bipolar junction transistor (BJT) operating in the linear region can be used as a series-pass element whose base current can be controlled to change the resistance presented to the input source. Figure 4.6(a) shows the basic schematic of the series regulator with the series-pass component being a transistor Q.

Figure 4.6(b) shows the series regulator wherein the base drive for the series-pass transistor Q is being controlled by the operational amplifier (op-amp). The output voltage V_o is attenuated by two resistors R_1 and R_2 and fed to the "−" terminal of the op-amp. The "+" of the op-amp is fed from a Zener voltage reference. The output of the op-amp drives the base of the transistor Q. If the output voltage increases, the "−" terminal of the op-amp will increase. As the "+" terminal is at a constant reference potential, the output of the op-amp will decrease the drive to the base of the transistor Q. This will increase the resistance presented

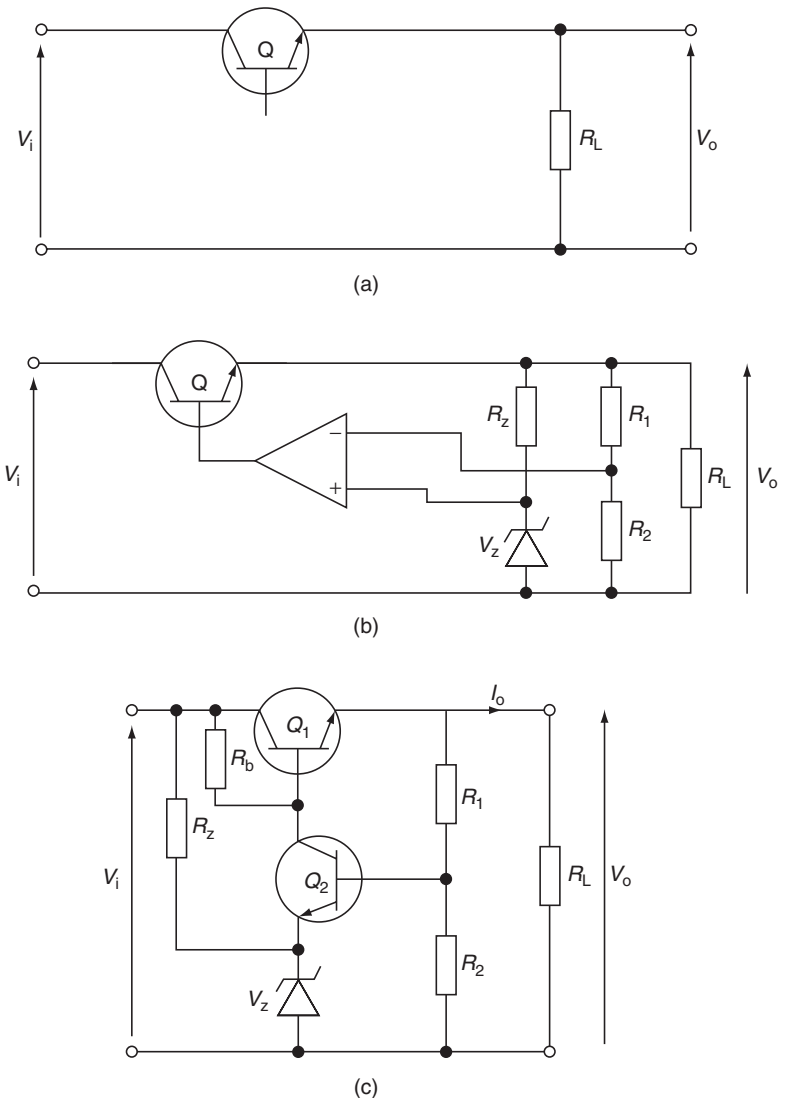


Figure 4.6 Series regulator: (a) Series regulator using transistor as the series pass; (b) the series-pass transistor is controlled by an op-amp; (c) the series pass is controlled by another transistor amplifier.

by Q to the input and thereby increase the drop across the collector to emitter of the transistor Q. As a consequence, the output voltage will decrease and be brought back to the equilibrium state. All these events occur simultaneously at an instant. Likewise if the output voltage were to decrease, the drive to Q will increase thereby decreasing the drop across the collector–emitter of Q. This will tend to increase the output voltage and bring it back to the equilibrium state.

The circuit of Figure 4.6(c) is similar to that of Figure 4.6(b) except that the op-amp is replaced by a transistor (BJT) amplifier Q_2 as shown. The Zener voltage reference is fed to the emitter terminal of Q_2 and the attenuated output voltage is fed to the base terminal of Q_2. The transistor Q_1 acts as the series-pass transistor. The base of the transistor Q_1 is at potential $(V_o + V_{be1})$. Thus, the current through the bias resistor R_b is

$$I_{Rb} = \frac{V_i - V_o - V_{be1}}{R_b} \tag{4.15}$$

I_{Rb} can be considered as a current source that diverts the current between the base of Q_1 and the collector of Q_2 depending on the output voltage.

If the output voltage decreases, then the base–emitter potential (V_{be2}) of Q_2 decreases. This will reduce the collector-current flow of Q_2. As a consequence, the portion of I_{Rb} that will be diverted to the base of Q_1 will increase, resulting in reduced collector–emitter resistance of Q_1. This will decrease the drop across the collector–emitter of Q_1 and will make the output voltage to increase and be brought back to the equilibrium state. In a similar manner, the output voltage regulation will occur when the output voltage increases.

The resistances R_1 and R_2 are chosen based on the following governing equation:

$$V_o\left(\frac{R_2}{R_1 + R_2}\right) = V_z + V_{be2} \tag{4.16}$$

Allow a current of about 1% of the full load current to flow through R_1 and R_2. This constraint and Eq. (4.16) can be used to decide the values of R_1 and R_2. R_z is chosen such that the current flowing through the Zener diode biases it in the breakdown region of the static characteristic curve. R_b is chosen based on Eq. (4.15) with the nominal value of the input voltage V_i.

Negative Voltage Regulator

In the series regulator of Figure 4.6, the negative rail is used as the circuit ground. However, if one has to use the positive rail as the circuit ground, then the schematic is modified as shown in Figure 4.7. The positive output voltage series regulators have the series-pass component in the positive rail and the voltages are measured with reference to the negative rail as indicated in Figure 4.7(a). The NPN series-pass transistor Q is replaced by an equivalent PNP series-pass transistor as shown in Figure 4.7(b). In this figure, the series-pass element is included in the negative rail and the voltages are measured with respect to the positive rail. Therefore the voltages are negative in the circuit of Figure 4.7(b).

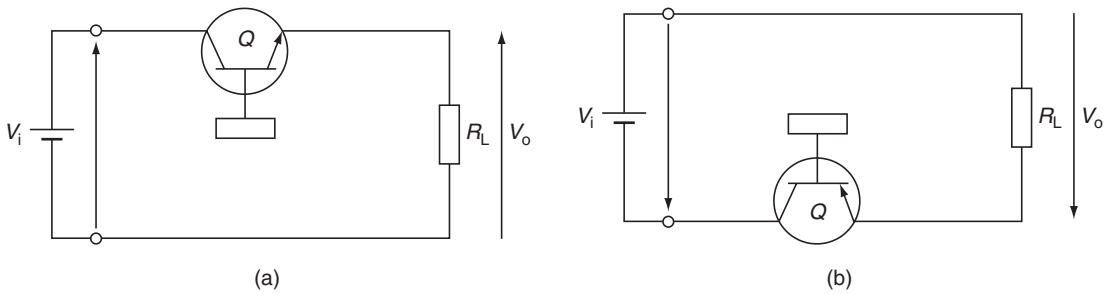


Figure 4.7 Series regulator with (a) negative rail as circuit ground; (b) positive rail as circuit ground.

The negative regulator of Figure 4.7(b) is implemented in Figure 4.8(a). The schematic of the negative output series regulator shown in Figure 4.8(a) is similar to the positive output regulator of Figure 4.6 except that the transistors are replaced by PNP transistors to handle reverse current flow direction. The Zener voltage reference is also reverse as indicated.

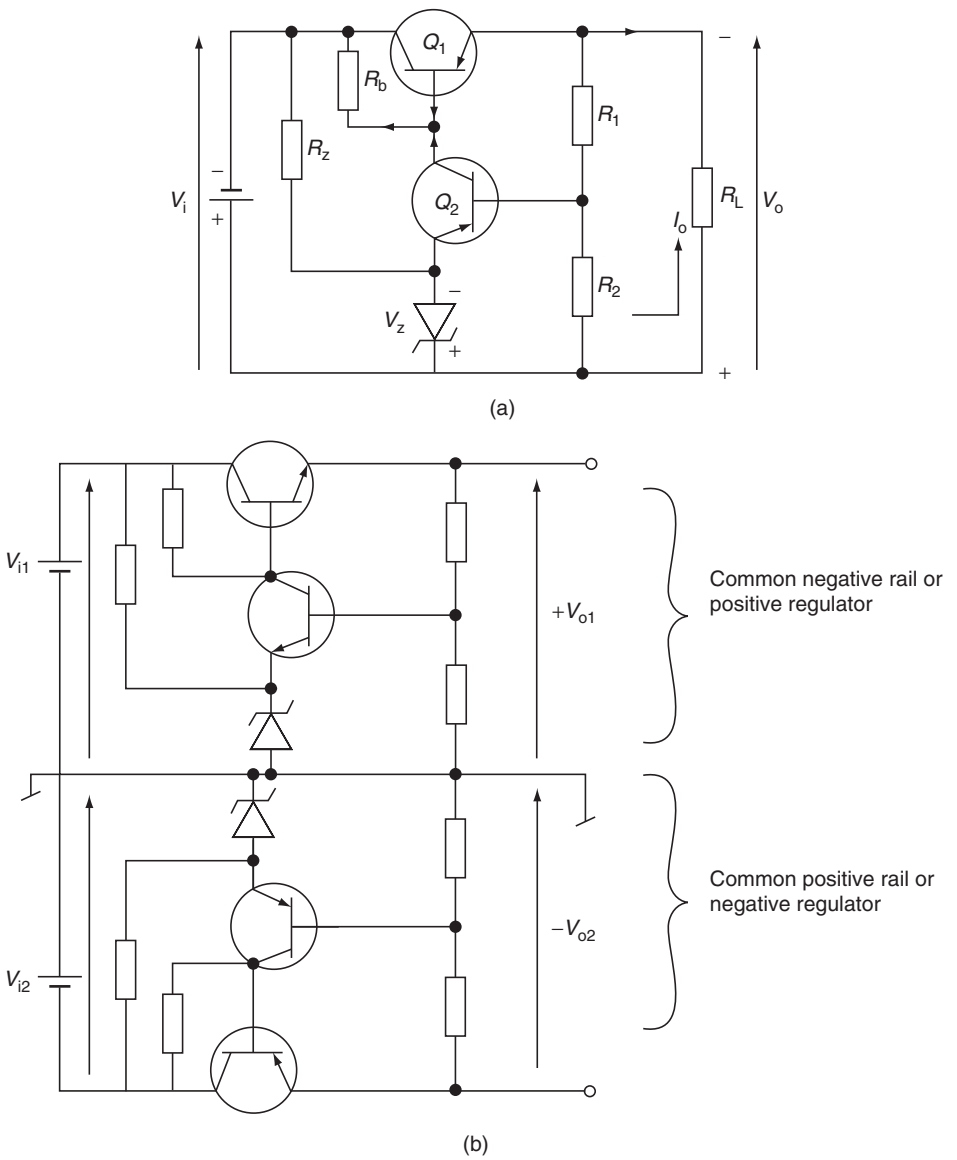


Figure 4.8 (a) Series regulator with negative output voltage; (b) dual power supply linear regulator.

The operation of the negative output voltage regulator of Figure 4.8(a) is also very similar to that of the positive voltage regulator of Figure 4.6. Here too, the current through R_b can be assumed to be a current source as the base of Q_1 is at potential $-V_o - V_{be1}$ and the other end of R_b is at $-V_i$ potential. Thus,

$$I_{Rb} = \frac{V_o + V_{be1} - V_i}{R_b} \tag{4.17}$$

If the output voltage becomes more negative, then the potential at the base of Q_2 decreases as compared to the emitter of Q_2. This increases the drive for the PNP transistor and the contribution of collector current of Q_2 to I_{Rc} increases, thereby decreasing the contribution of the base current of Q_1 to I_{Rb}. As this decreases the base drive for Q_1, the emitter–collector voltage drop of Q_1 increases and brings V_o to the equilibrium state. Similar regulation action occurs when Vo becomes less negative.

The positive rail regulators and the negative rail regulators can be used together to obtain both positive and negative power supplies with a common ground. A typical dual supply linear regulator circuit is shown in Figure 4.8(b).

IC Linear Regulators

The linear regulators are available commercially in the form of integrated circuits (IC). The IC regulators can be classified into three major categories, namely,

1. fixed regulators;
2. variable regulators;
3. variable regulators with current boost.

The fixed regulator types are available as three-pin regulators. The more popular among them are the 78xx and 79xx series. Here xx represents the output voltage of the regulator. For example, 7805 is a 5 V positive voltage regulator, 7812 is a 12 V positive voltage regulator and so on. The 79xx series are negative voltage regulators that are complementary to the 78xx series.

The variable regulators are also available as three-pin regulators with three pins, viz. input, output and adjust. The "adjust" pin is used for varying and setting the output voltage of the regulator. The 317 regulator is a popular variable voltage regulator. The 350 is also another three-pin adjustable regulator like the 317 but with higher load current capability. The 337 is a negative voltage regulator that is complementary to the 317 positive voltage regulator.

There is a class of IC regulators that gives flexibility both in output voltage setting and also in the current-carrying capability by giving facility to include an external boost transistor. The 723 is a very popular linear regulator IC in this class.

Example: 317 IC Regulator

The schematic of the 317 IC linear regulator is shown in Figure 4.9. The characteristic feature of this regulator is the voltage across the output pin (OUT) and the adjust pin (ADJ). The voltage between these two pins is maintained at 1.25 V for this regulator. The output voltage V_o is decided by the ratio of R_1 and R_2. Referring to Figure 4.9, the output voltage is given as

$$V_o = 1.25\left(1 + \frac{R_2}{R_1}\right) \tag{4.18}$$

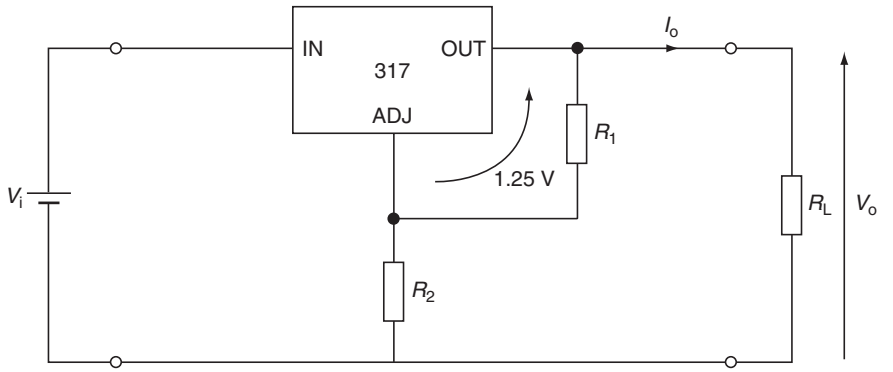


Figure 4.9 | Basic schematic of 317 IC linear regulator.

R_1 is recommended by the manufacturer to be around 240 Ω. R_2 can be calculated from the output voltage requirement and Eq. (4.18). The maximum input voltage that can be applied is 40 V. The 317 IC regulator needs a minimum differential voltage of 3 V between its input and output pins. Thus the maximum output voltage can be 37 V. The 317 IC is rated to handle 1.5 A and the 350 IC is rated to handle 3 A. All other aspects of 350 IC are similar to the 317 IC.

The circuit of Figure 4.9 is sufficient for the basic operation of the regulator. However there are a few components that need to be included to improve the reliability of the circuit. The complete circuit schematic of the 317 IC regulator circuit is shown in Figure 4.10.

The regulation of the 317 IC is best when the input and the output load are pure DC. However, there will be ripples on the input side and the load may be switching loads like digital circuit loads. In such cases there will be AC components present. The regulator performance deteriorates for the AC components. Therefore, the AC components in the load side that are generated due to digital loads should not be passed

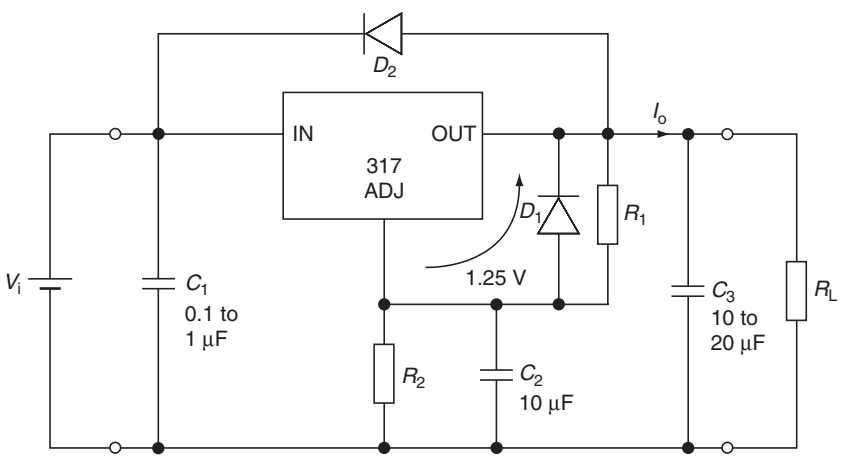


Figure 4.10 | 317 IC linear regulator complete circuit schematic.

through the regulator to the input side. Likewise the AC ripples present in the input side should not be passed to the load side through the regulator. The AC components should be bypassed by using capacitors that will provide low impedance paths for the AC components on either side. C_1 and C_3 are the input side and the output side AC bypass capacitors, respectively. A capacitor C_2 is connected between the ADJ pin of the regulator and the circuit ground. This capacitor acts as a buffer to stabilize the fluctuations in the voltage at the ADJ pin. This provides increased input ripple rejection. However, during turn OFF of the power supply, the output capacitor C_3 will discharge its charge to the load. The capacitor C_2 will try to discharge through the regulator. However, a large discharge current through the ADJ pin will damage the 317 IC. To provide an alternate discharge path for the capacitor C_2, diode D_1 has been introduced. Capacitor C_2 can discharge through D_1 into the output load. Diode D_2 is a protection diode. During situations when the input voltage suddenly becomes zero during power outage, the input output voltage differential across the 317 IC will become negative and will be equal to $-V_o$. This will damage the regulator IC. To prevent this, D_2 will ensure that the drop across the input and output of the regulator IC will be clamped to −0.7 V if ever the output voltage increases beyond the input voltage thereby protecting the regulator IC.

Practical Tips

One of the main benefits of the linear regulators is the high quality of performance. However, there are integration issues that can have adverse effects on the performance of the regulation. One such crucial issue is the layout of the input side capacitor. The input side capacitor carries high peak currents as shown and also discussed in the previous chapter. If the capacitor is laid out as shown in Figure 4.11(b), then the output voltage is given as

$$V_o = V_{reg} + i_c R_{lead}$$

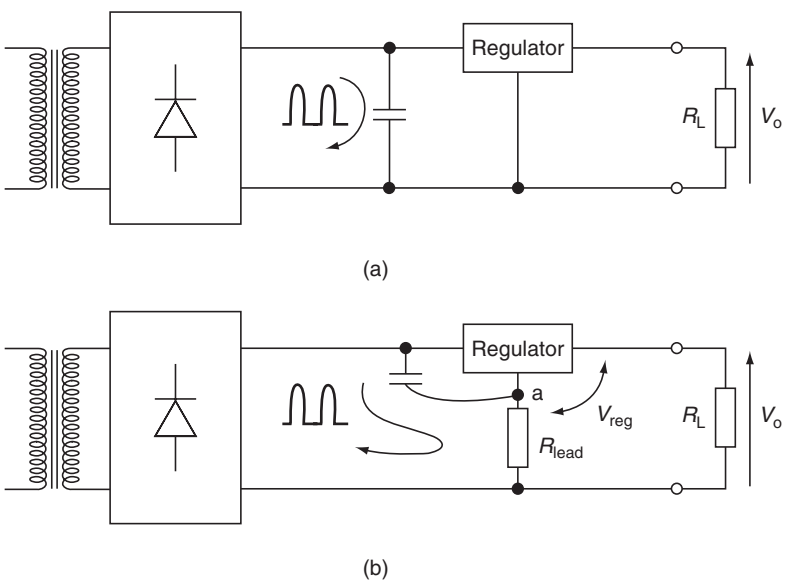


Figure 4.11 Loss of regulation due to incorrect layout of input capacitor (a) correct; (b) incorrect.

The term $i_c R_{lead}$ in the output voltage equation will cause de-regulation with load. Therefore care should be taken to layout the input side capacitor such that it does not interfere with the regulation. The layout of Figure 4.11(a) is correct and will not deteriorate the output regulation.

Another important layout issue that affects the output regulation is shown in Figure 4.12 with respect to the 317 IC regulators. In Figure 4.12(a), the lead resistance R_{lead} comes within the regulation loop of the IC regulator and in Figure 4.12(b), the layout is such that the lead resistance through which the load current flows is out of the regulation loop of the IC regulator. For the layout of Figure 4.12(a), the output is given as

$$V_o = \left(1.25 - I_o R_{lead}\right)\left(1 + \frac{R_2}{R_1}\right) = 1.25\left(1 + \frac{R_2}{R_1}\right) - I_o R_{lead}\left(1 + \frac{R_2}{R_1}\right) \tag{4.19}$$

For the layout of Figure 4.12(b), the output is given as

$$V_o = 1.25\left(1 + \frac{R_2}{R_1}\right) - I_o R_{lead} \tag{4.20}$$

From Eqs. (4.19) and (4.20) it is evident that in the case of layout of Figure 4.12(b), the term affecting regulation is only $I_o R_{lead}$, whereas the term affecting regulation in the case of layout of Figure 4.12(a) is

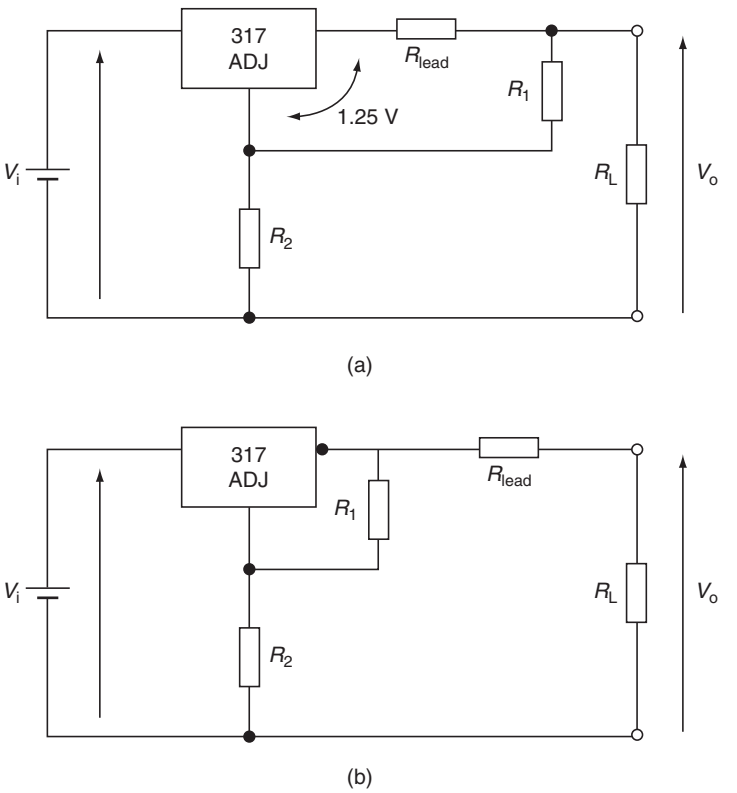


Figure 4.12 | Effect of lead resistance in 317 IC or any variable IC regulator.

$I_o R_{lead}[1+(R_2/R_1)]$ which is higher than that for Figure 4.12(b). Therefore, the layout of Figure 4.12(b) is the preferred layout.

Four-Wire Connection

Another important consideration in the linear regulators is the issue of high current loads. In the case of high current loads, especially if they are located remotely, the lead resistance R_{lead} (Figure 4.13) will significantly affect the output voltage regulation. Referring to the circuit schematic shown in 4.13, V_{reg} is the voltage that is regulated by the regulator. However, as the actual load is located remotely, the actual output voltage is across the load resistor R_L and is given as

$$V_o = V_{reg} - I_o R_{lead} \tag{4.21}$$

From Eq. (4.21), it is evident that the actual output voltage is different from the regulated voltage V_{reg} by a load-dependent term $I_o R_{lead}$. This will make the output voltage regulation for changes in the load poorer.

A solution for this is the four-wire connection as shown in Figure 4.14. The regulator provides four output terminals 1–4 as indicated in Figure 4.14, viz., (a) positive rail or + power terminal; (b) positive

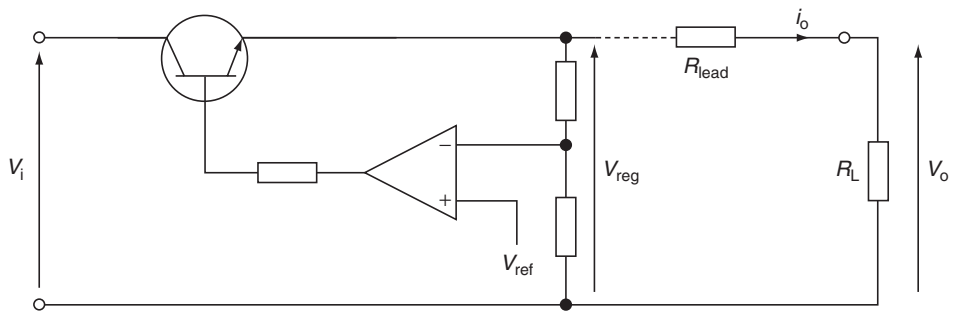


Figure 4.13 | Effect of lead resistance in high current regulators.

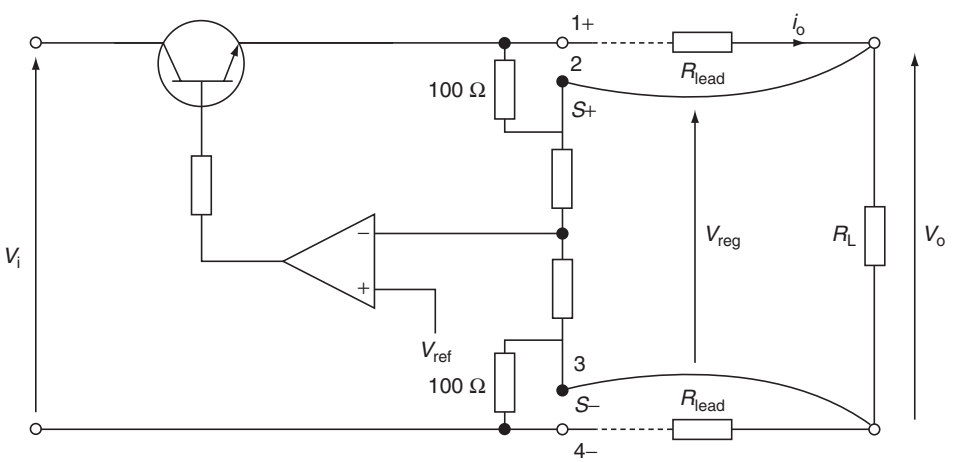


Figure 4.14 | Four-wire connection for high current linear regulators.

of the output voltage sense or S+ terminal; (c) negative of the output voltage sense or S− terminal and (d) negative rail or − power terminal.

The load current flows through the positive rail terminal, through the load and finally returns through the negative rail terminal. The actual output voltage V_o is sensed by the sensing leads that are connected to terminals 2 and 3. The currents through the sensing leads are very small as compared to the load current I_o. Therefore, the sense terminals will essentially measure the actual remote output voltage V_o. In this way, even the drops across the lead resistances in the positive rail and negative rail are included inside the regulation loop, thereby regulating the actual output V_o. Sometimes terminals 1 and 2 and terminals 3 and 4 are connected with resistances of about 100 Ω as shown in Figure 4.14. This will enable either two-wire connections or four-wire connections to be used depending on the remoteness of the load, wiring layout and lead resistances.

Protection

Protection circuitry is an integral part of any system for improving the availability of the system. The linear regulator should, in general, be protected against the following:

1. over current;
2. over voltage;
3. reverse voltages;
4. reverse voltage drops across the series component.

For the shunt linear regulators, the series component is a fixed resistor. The power rating of the resistor is normally chosen to be

$$P_{\mathrm{Rs}} = \frac{(V_{\mathrm{i\text{-}max}} - V_{\mathrm{z}})^2}{R_{\mathrm{s}}}$$

However, if the output gets short circuited, then the entire output voltage will drop across the series resistor. The dissipation will exceed the rated power dissipation and damage the series component. To take care of this extreme situation, the series resistor R_s in the case of the shunt regulator should have a power rating of

$$P_{\mathrm{Rs}} = \frac{V_{\mathrm{i\text{-}max}}^2}{R_{\mathrm{s}}}$$

For series regulators, Figure 4.15 shows a very simple and effective current protection scheme. Two components, transistor Q_s and current resistor R_s are used for current protection. The current sense resistor is chosen such that when the load current limit is reached, the voltage drop across it is 0.5 V. Any further increase in the load current will increase the drop across R_s which in turn increases the base–emitter voltage of Q_s. This increased base–emitter voltage will make Q_s to draw increased collector current. As the current through R_b is a current source, the increase in the collector current drawn by Q_s is at the expense of the base current of Q_1. Thus, as the drive to Q_1 decreases, Q_1 presents a higher resistance to the input source and the drop across Q_1 increases thereby decreasing the output voltage and load current. In this manner, the combination of R_s–Q_s will provide over current limiting in linear series regulators.

The over voltage protection is mainly to ensure that the load is not damaged. The over voltage may occur if the series pass of the linear regulator is shorted or due to an electrostatic coupling at the output. The over voltage protection is implemented by means of the crow-bar protection circuit as shown in Figure 4.16. Under normal operating conditions, (a) the thyristor, Th_1, is OFF, (b) the Zener diode D_{cb} is reverse-biased and in OFF region and (c) the potential of the thyristor gate is zero. If for some reason the output voltage goes beyond a particular limit, the Zener diode D_{cb} breaks down and a potential of $[(V_o - V_{\mathrm{z\text{-}cb}}) / (R_{1b} + R_{2b})]R_{2b}$

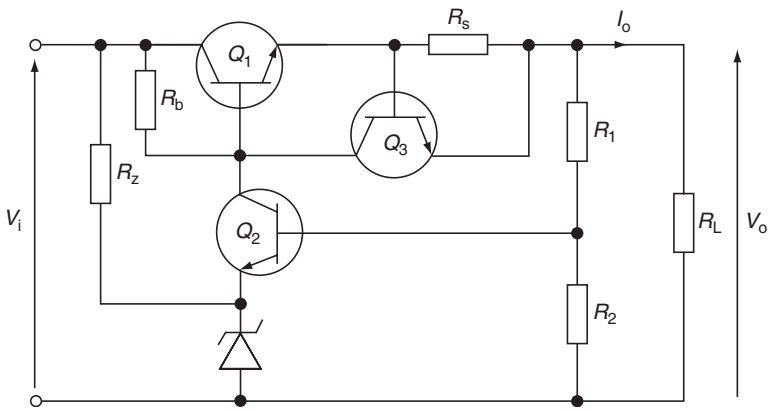


Figure 4.15 | Current protection.

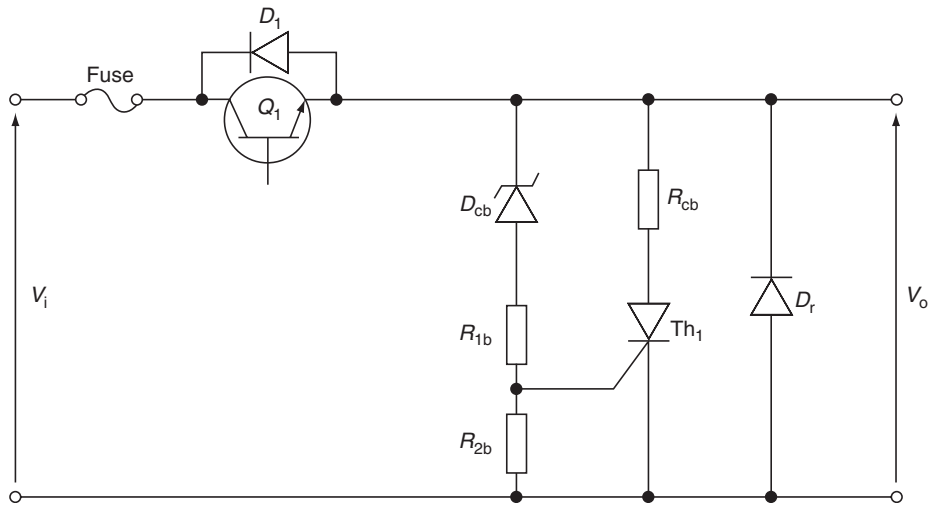


Figure 4.16 | Voltage protections.

is applied at the gate of the thyristor Th_1. The thyristor turns ON and connects a small value resistance R_{cb} across the output V_o, effectively shorting the output with R_{cb}. This will bring down the output voltage within safe limits. The thyristor switches OFF only after its current is brought below the holding current value. For crowbar protection, a fuse is included in series with the input source. When the output is short circuited due to the action of the Th_1, the current from the input source increases and blows the fuse. This will bring the current in the thyristor below the holding current value and switches OFF the thyristor Th_1. Alternately, this is done by switching OFF the input source V_i by turning OFF a series switch that can be included in series with the input source or the current protection scheme of Figure 4.15 can be used. If Th_1 turns ON due to over voltage condition, the current through the sense resistor R_s of Figure 4.15 will increase and clamp at 0.7 V. Transistor Q_s will be in ON-state and divert all the current from I_{Rb} to its collector. This

will cut-off Q_1 which will bring down the thyristor current below the holding value and turn OFF the thyristor. However, it should be noted that if the output over voltage has occurred due to Q_1 collector–emitter short, then the current limit circuit will not work. In such cases either the fuse in series with the input will blow or the collector–emitter of Q_1 that was shorted will open.

Diode D_r is connected as shown in the figure to prevent any negative voltage being applied to the load. Similar reverse diode may be connected at the input side to prevent any accidental reverse voltage being applied to the positive rail regulator, thereby preventing any damage to the linear regulator and the load. Diode D_1 is used to prevent any negative voltage drop occurring across the series-pass element. This can happen when the input is turning OFF and the output capacitor is still holding charge. Under these conditions, diode D_1 will conduct and clamp the voltage across the series-pass element to −0.7 V.

Current Regulation

Till now, the linear regulators for output voltage regulation are discussed. However, similar principles can be used for regulating the output current to obtain a current source. The main principle in a current regulator is shown in Figure 4.17(a). Here the voltage across a resistor V_R is held constant by means of a voltage regulator that may be a shunt or a series regulator. This makes the current V_R/R through that resistor constant which is directed to the output as the load current.

Figure 4.17(b) shows the circuit implementation of the current regulator. A modified shunt regulator with a PNP transistor is used in the circuit schematic of the current regulator. The voltage across the emitter resistor R_e is given as $V_z - V_{be}$. The emitter current is therefore given as

$$I_{Re} = \frac{V_z - V_{be}}{R_e} \tag{4.22}$$

From Eq. (4.22), it can be observed that V_z and V_{be} are constants and hence the current through R_e is a constant. For a high β transistor, the current through the collector of the transistor Q is almost the same as I_{Re}. This current flows through the connected load resistor R_L as a constant current developing an output voltage of

$$V_o = I_L R_L = \left(\frac{V_z - V_{be}}{R_e} \right) R_L \tag{4.23}$$

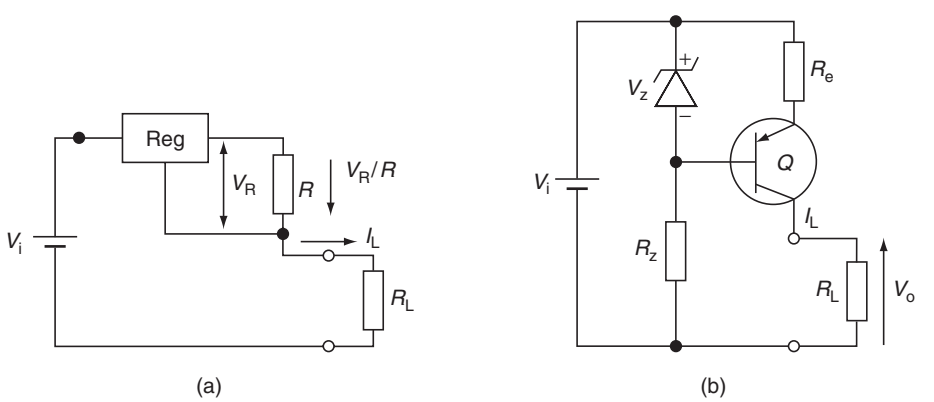


Figure 4.17 | (a) Current regulation principle; (b) current regulator circuit schematic.

The maximum output voltage that can develop is given by

$$V_{\text{o-max}} = V_{\text{i-max}} - V_{\text{z}} + V_{\text{be}} \tag{4.24}$$

4.3 Parameters of the Linear Regulator

The following four parameters of the linear regulator determine the quality of the regulator:

1. efficiency;
2. line regulation;
3. load regulation;
4. temperature regulation.

The first parameter is related to the power delivered to the output with respect to the power consumed from the source. This primarily affects the volumetric power density of the linear regulator. The power dissipated in the series-pass element determines the size of the heat sink that needs to be employed to ensure proper heat flow to the ambient. The remaining three parameters provide quantitative measures for the performance quality of the output voltage. This performance quality of the linear regulator is measured by regulation of the output voltage for (a) variations in the line, (b) variations in load and (c) variations in temperature. This is mathematically represented as

$$\Delta V_{\text{o}} = \frac{\partial V_{\text{o}}}{\partial V_{\text{i}}} \Delta V_{\text{i}} + \frac{\partial V_{\text{o}}}{\partial I_{\text{o}}} \Delta I_{\text{o}} + \frac{\partial V_{\text{o}}}{\partial T} \Delta T \tag{4.25}$$

where $\partial V_{\text{o}} / \partial V_{\text{i}}$ is the line-regulation coefficient with I_{o} and T constant; $\partial V_{\text{o}} / \partial I_{\text{o}}$ the load-regulation coefficient with V_{i} and T constant; $\partial V_{\text{o}} / \partial T$ the temperature-regulation coefficient with V_{i} and I_{o} constant. For any linear regulator selection, these four important parameters are essential and need to be quantified and specified. The analysis of the regulators will also focus on these parameters.

4.4 Analysis of a Series Regulator

Till now, the regulator circuits were discussed in an operational sense wherein one could design and implement the regulator circuits. However, to obtain greater insight into the operation of the regulator so that one may improve the circuits, a more systematic method of analysis has to be performed. In this section, a systematic analysis will be performed on a series linear regulator circuit (shown in Figure 4.18). The analysis will bring out the problems in this topology and will also suggest remedial measures that can be taken to address the issues. Without loss of generality, a similar approach may be used to analyze other linear regulator circuits.

First-Level Modeling

To analyze the regulator shown in Figure 4.18, the transistor and the Zener diode are replaced by the first-level idealized model. The first-level model for the transistor and the Zener diode are shown in Figure 4.19. For the NPN transistor shown in Figure 4.19(a), the first-level model is shown in Figure 4.19(b). The collector current is a dependent current source of value βi_{b}. The first-level model of a Zener diode is a DC source of value V_{z} as shown in Figure 4.19(d).

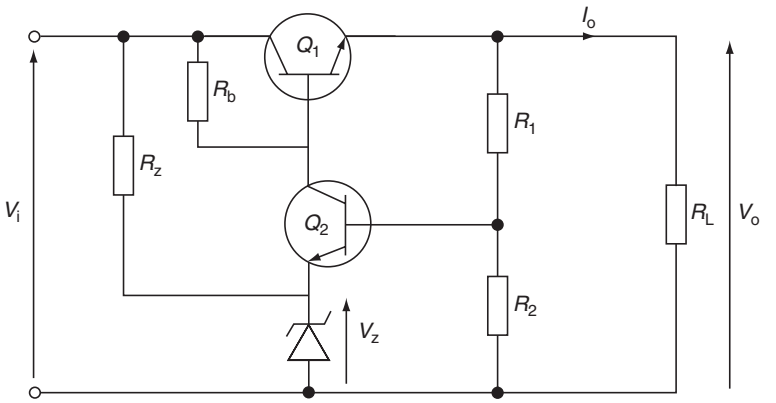


Figure 4.18 | Linear series regulator.

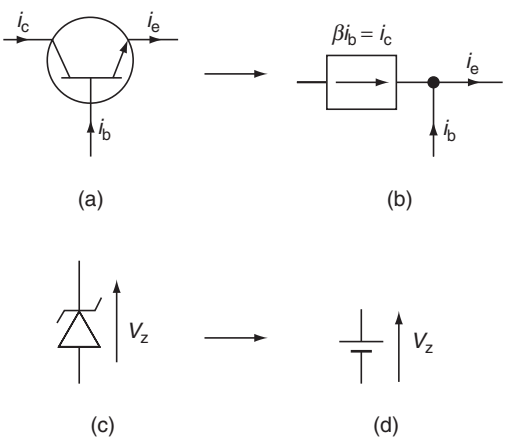


Figure 4.19 | (a) NPN transistor; (b) NPN transistor first-level model; (c) Zener diode; (d) first-level model of Zener diode.

Referring to Figure 4.18, both the transistors Q_1 and Q_2 are replaced by their first-level model. Similarly, the Zener diode is also replaced by its first-level idealized model. The resulting equivalent circuit is shown in Figure 4.20.

By applying Thevenin's theorem to the base portion of the transistor Q_2 of Figure 4.20, the modified equivalent circuit representation of the series regulator is obtained as shown in Figure 4.21.
Referring to Figure 4.21, the base current of Q_2 is given as

$$i_{b2} = \frac{V_o[R_2/(R_1+R_2)] - V_z}{R_1 /\!/ R_2} = \frac{V_o}{R_1} - V_z\left(\frac{R_1+R_2}{R_1R_2}\right) \tag{4.26}$$

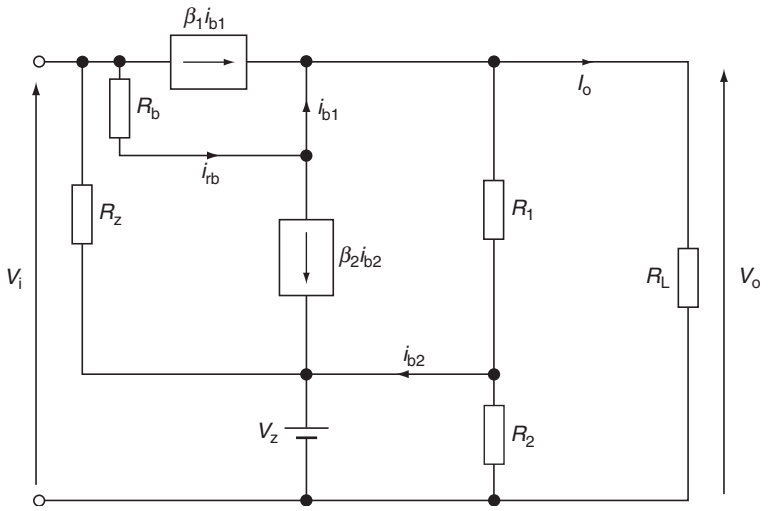


Figure 4.20 Equivalent circuit of the series regulator using the first-level models of the transistor and the Zener diode.

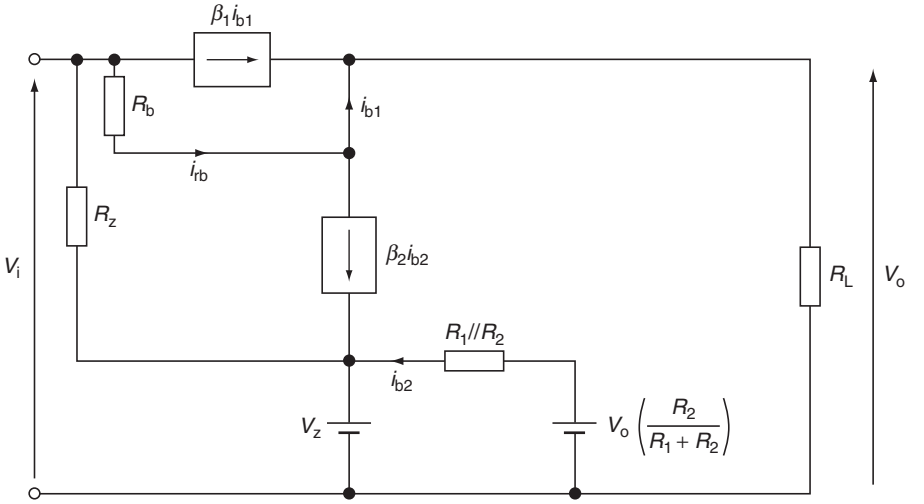


Figure 4.21 Equivalent circuit of the series regulator.

The current through R_b is given as

$$i_{rb} = \frac{V_i - V_o}{R_b} \tag{4.27}$$

$$\begin{aligned} i_{b1} &= i_{rb} - \beta_2 i_{b2} \\ &= \frac{V_i - V_o}{R_b} - \left\{ \beta_2 \frac{V_o}{R_1} - \beta_2 V_z \left(\frac{R_1 + R_2}{R_1 R_2} \right) \right\} \end{aligned} \tag{4.28}$$

The regulating action of the series regulator is defined by Eq. (4.26). If the output voltage V_o increases, then i_{b2} increases. An increase in i_{b2} will cause a decrease in i_{b1} as seen from Eq. (4.28). This in turn will reduce the base drive for Q_1 which will present a higher resistance to the input source. As a consequence, the collector–emitter of Q_1 will drop more voltage and the output voltage will reduce and return back to the equilibrium state. A similar regulating action occurs when the output voltage V_o decreases.
The output voltage is given as

$$\begin{aligned} V_o &= (i_{b1} + \beta_1 i_{b1}) R_L = (\beta_1 + 1) i_{b1} R_L \\ &= (\beta_1 + 1) R_L \left[\frac{V_i - V_o}{R_b} - \beta_2 \frac{V_o}{R_1} + \beta_2 V_z \left(\frac{R_1 + R_2}{R_1 R_2} \right) \right] \end{aligned} \tag{4.29}$$

The output relationship as given in Eq. (4.29) is re-arranged such that it is of the form

$$V_o = k_1 V_i + k_2 V_z \tag{4.30}$$

where

$$k_1 = \frac{(\beta_1 + 1) R_L / R_b}{[(\beta_1 + 1) R_L / R_b] + [(\beta_1 + 1) \beta_2 (R_L / R_1)] + 1}$$

and

$$k_2 = \frac{(\beta_1 + 1) \beta_2 R_L [(R_1 + R_2) / R_1 R_2]}{[(\beta_1 + 1) R_L / R_b] + [(\beta_1 + 1) \beta_2 (R_L / R_1)] + 1}$$

It should be noted that in the above derivation for V_o, the load current I_o is taken as $(\beta_1 + 1) i_{b1}$. However, I_o is actually $(\beta_1 + 1) i_{b1} - i_{R1}$ where i_{R1} is the current through R_1. Referring to Figure 4.20, this is given as $(V_o - V_z) / R_1$. As R_1 is large, i_{R1} is neglected in the above first-level derivation for V_o. However, in later discussions, the contribution of i_{R1} to the load current is also considered.

Referring to Eq. (4.30), there are a few conclusions that one can draw at this point of the analysis. If V_o is to be independent of the input source V_i, then k_1 should be zero and k_2 should be a constant. The factor k_1 can be written as

$$k_1 = \frac{1}{1 + \beta_2 (R_b / R_1) + [R_b / (\beta_1 + 1) R_L]}$$

In this equation, only the term $\beta_2 (R_b / R_1)$ is greater than unity. If the transistor Q_2 is chosen as a high gain transistor, then β_2 will be high. In fact, β_2 will be much higher than the ratio of R_b/R_1. As a consequence, this term will be much greater than 1 and thus

$$k_1 \ll 1 \tag{4.31}$$

The factor k_2 in Eq. (4.30) can be written as

$$k_2 = \frac{\beta_2 [(R_1 + R_2) / R_1 R_2]}{(1 / R_b) + \beta_2 (1 / R_1) + [1 / (\beta_1 + 1) R_L]}$$

Due to the β_2 terms both in the numerator and the denominator, the value of k_2 is significant and

$$k_2 \gg k_1 \tag{4.32}$$

From the inequalities in Eqs. (4.31) and (4.32), one can say that the term $k_1 V_i$ is negligibly small. Thus

$$V_o \approx k_2 V_z$$

which is a regulated voltage.

Effect of Current through R_b

From Figure 4.21, it can be observed that the current through the base of the series-pass transistor Q_1 is

$$i_{b1} = i_{rb} - \beta_2 i_{b2}$$

where

$$i_{b2} = \frac{V_o}{R_1} - V_z\left(\frac{R_1 + R_2}{R_1 R_2}\right)$$

If R_1, R_2, V_z and V_o are constants, then i_{b2} will be a constant. For a given temperature, β_1 can also be considered to be a constant. Therefore, the term $\beta_2 i_{b2}$ in the i_{b1} equation is a constant and does not contribute to loss of regulation. The other term in the i_{b1} equation is the current through R_b. This is given as

$$i_{rb} = \frac{V_i - V_o}{R_b} \tag{4.33}$$

In Eq. (4.33), one can observe that i_{rb} is a function of both V_i and V_o. If V_i varies, i_{rb} also varies and hence i_{b1} varies which in turn causes V_o to vary as

$$V_o = (\beta_1 + 1) i_{b1} R_L \tag{4.34}$$

Therefore from Eqs. (4.33) and (4.34) it is evident that the loss of output regulation is due to i_{rb} not being a constant in the face of variations in V_i.

The solution to this is as follows: It is required that i_{rb} is constant and independent of variations in the input source voltage V_i. If this is achieved then i_{b1} and hence V_o will become independent of variations in the input source voltage V_i. One possible method to make i_{rb} independent of variations in V_i is to replace the resistor R_b by a constant-current source as illustrated in Figure 4.17(b). The constant-current source circuit is integrated into the series regulator circuit and the resulting modified circuit is depicted in Figure 4.22.

The operation of the constant-current source portion is as discussed in the section "Current Regulation". The current i_{rb} is now a constant that is independent of variations in the input voltage V_i. Now the output voltage V_o is given as

$$\begin{aligned} V_o &= (i_{b1} + \beta_1 i_{b1}) R_L = (\beta_1 + 1) i_{b1} R_L \\ &= (\beta_1 + 1) R_L \left[i_{rb} - \beta_2 \frac{V_o}{R_1} + \beta_2 V_z \left(\frac{R_1 + R_2}{R_1 R_2}\right)\right] \end{aligned} \tag{4.35}$$

From Eq. (4.35), it can be observed that as i_{rb} is a constant, V_o is essentially independent of V_i and depends only on V_z.

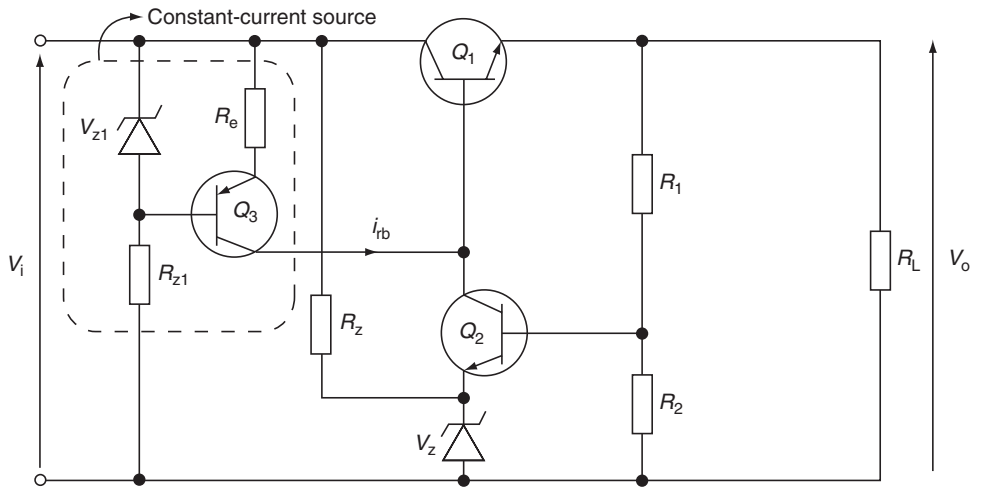


Figure 4.22 | Series regulator circuit: R_b is replaced by a constant-current source.

Effect of Current through R_1

In the discussions till now, the output load current I_o is considered to be $(\beta_1 + 1)i_{b1}$ wherein the current through R_1 is neglected. If the current through R_1 is also considered in the output voltage equation with respect to the series regulator equivalent circuit of Figure 4.21, then V_o is given as

$$\begin{aligned} V_o &= [(\beta_1 + 1)i_{b1} - i_{R1}]R_L \\ &= \left[(\beta_1 + 1) i_{b1} - \left(\frac{V_o - V_z}{R_1} \right) \right] R_L \end{aligned} \tag{4.36}$$

Substituting for i_{b1} from Eq. (4.28) into Eq. (4.36), one obtains

$$V_o = (\beta_1 + 1)R_L \left[\frac{V_i - V_o}{R_b} - \beta_2 \frac{V_o}{R_1} + \beta_2 V_z \left(\frac{R_1 + R_2}{R_1 R_2} \right) \right] - \left(\frac{V_o - V_z}{R_1} \right) R_L$$

The above equation for V_o can be re-arranged in the following form:

$$V_o = k_1 V_i + k_2 V_z \tag{4.37}$$

where

$$k_1 = \frac{(\beta_1 + 1)R_L / R_b}{(\beta_1 + 1)(R_L / R_b) + (\beta_1 + 1)\beta_2 (R_L / R_1) + (R_L / R_1) + 1}$$

and

$$k_2 = \frac{(\beta_1 + 1)\beta_2 R_L [(R_1 + R_2) / R_1 R_2] + (R_L / R_1)}{(\beta_1 + 1)(R_L / R_b) + (\beta_1 + 1)\beta_2 (R_L / R_1) + (R_L / R_1) + 1}$$

The load current I_o is given as

$$I_o = \frac{V_o}{R_L} = (\beta_1 + 1)\left[\frac{V_i - V_o}{R_b} - \beta_2 \frac{V_o}{R_1} + \beta_2 V_z\left(\frac{R_1 + R_2}{R_1 R_2}\right)\right] - \left(\frac{V_o - V_z}{R_1}\right)$$

which can be re-arranged as

$$I_o = \frac{(\beta_1 + 1)}{R_b} V_i - \left(\frac{(\beta_1 + 1)}{R_b} + \frac{(\beta_1 + 1)\beta_2}{R_1} - \frac{1}{R_1}\right) V_o + \left[(\beta_1 + 1)\beta_2\left(\frac{R_1 + R_2}{R_1 R_2}\right) + \frac{1}{R_1}\right] V_z \tag{4.38}$$

Regulator Parameters

The regulation parameters, viz., line-regulation coefficient, load-regulation coefficient and temperature-regulation coefficient can be estimated based on the equations for V_o and I_o as given in Eqs. (4.37) and (4.38).

Line-Regulation Coefficient

The line-regulation coefficient can be obtained from Eq. (4.37). It is the variation of the output voltage V_o for variations of the input voltage V_i under the constraints of constant load and temperature. It is given as

$$\frac{\partial V_o}{\partial V_i} = \frac{(\beta_1 + 1) R_L / R_b}{(\beta_1 + 1)(R_L / R_b) + (\beta_1 + 1)\beta_2 (R_L / R_1) + (R_L / R_1) + 1} \tag{4.39}$$

In the case of the modified series regulator circuit wherein the resistor R_b is replaced by a constant-current source, i_{rb} is a constant. From Eq. (4.35), it can be observed that V_o is independent of V_i and therefore,

$$\frac{\partial V_o}{\partial V_i} = 0$$

Load-Regulation Coefficient

The load-regulation coefficient is a measure of the change in the output voltage V_o for changes in the load I_o under the constraints of constant temperature and input voltage V_i. The load-regulation coefficient can be obtained from Eq. (4.38). Thus,

$$\frac{\partial I_o}{\partial V_o} = -\left(\frac{\beta_1 + 1}{R_b} + \frac{(\beta_1 + 1)\beta_2}{R_1} - \frac{1}{R_1}\right)$$

The inverse of the above equation will give the load-regulation coefficient. This is given as

$$\frac{\partial V_o}{\partial I_o} = -\frac{R_b R_1}{(\beta_1 + 1) R_1 + (\beta_1 + 1)\beta_2 R_b - R_b} \tag{4.40}$$

Here $\partial V_o / \partial I_o$ is a measure of the output resistance R_o of the regulator circuit. The negative sign in Eq. (4.40) does not mean that R_o is negative, but it implies that the nature of the change in V_o with respective to changes in I_o is opposite in phase, that is, if I_o increases by ΔI_o, V_o decreases by ΔV_o and this is equal to $\Delta I_o R_o$.

From Eq. (4.40), it can be observed that the denominator contains a term that is a product of the βs, that is, the current gains of the transistors Q_1 and Q_2. This implies that R_o value is a very small value such that $R_o \ll 1$. If Q_1 and Q_2 are chosen such that their current gains are high, then R_o can be negligibly small.

In the case of the modified series regulator circuit wherein R_b is replaced by a constant-current source circuit, the load-regulation coefficient is given as

$$\frac{\partial V_o}{\partial I_o} = -\frac{R_1}{(\beta_1 + 1)\beta_2 - 1} \tag{4.41}$$

Temperature Effects

In the first-level modeling, as the transistors and the Zeners are considered as ideal. There is no contribution from these elements to the output voltage variations due to temperature changes. However, there will be changes in the output voltage for temperature changes due to variations in the value of the resistors with temperature. The effect of temperature variations on the output voltage is discussed later while considering non-ideal transistors and Zeners in the second-level model.

Problem 4.2

Consider a linear series regulation with $R_1 = 3.6$ kΩ, $R_2 = 2$ kΩ, $R_b = 3.6$ kΩ, $R_z = 20$ kΩ, $V_z = 6.55$ V, $R_L = 1$ kΩ, $V_i = 20$ V to 30 V, $V_{i\text{-nominal}} = 25$ V, $\beta_1 = \beta_2 = 100$. Calculate the output voltage a load change of 0.1 A to 0.2 A.

Solution

Applying the values to the variables in Eq. (4.37), one obtains

$$V_o = 9.89 \times 10^{-3} V_i + 2.77 V_z = 9.89 \times 10^{-3} V_i + 18.14$$

1. At $V_i = 25$ V, $V_o = 18.39$ V
2. At $V_i = 20$ V, $V_o = 18.3378$ V
3. At $V_i = 30$ V, $V_o = 18.4367$ V

For a change of 10 V in V_i, the change in output voltage is 0.0989 V.

$$R_o = \left|\frac{\partial V_o}{\partial I_o}\right| = -\frac{3600 \times 3600}{101 \times 3600 + 101 \times 100 \times 3600 - 3600} = 0.3529\ \Omega$$

For a load change of 0.1 A to 0.2 A,

$$\Delta V_o = 0.3529 \times 0.1 = 0.03529\ \text{V}$$

which is about 0.2% of V_o.

Range of Load Resistance R_L

Referring to the output voltage in Eq. (4.37), both k_1 and k_2 factors are functions of R_L. Therefore the output voltage V_o is not independent of R_L. Dividing the numerator and denominators of k_1 and k_2 by R_L, a term $1/R_L$ is present in the denominator of both k_1 and k_2. Referring to the denominators of k_1 and k_2 of the output voltage Eq. (4.37), if

$$\left\{(\beta_1+1)\frac{1}{R_b}+(\beta_1+1)\beta_2\frac{1}{R_1}+\frac{1}{R_1}\right\}\gg\frac{1}{R_L}$$

then the output voltage is not affected significantly by R_L. As R_L approaches infinity, that is, the series regulator approaches open circuit condition, $1/R_L$ anyway approaches zero and the above inequality constraint is easily met thereby making V_o independent of R_L. However, as R_L decreases, that is, the series regulator approaches the full-load or overload conditions, the dependency of V_o on R_L is more significant.

For Problem 4.2, wherein R_1 = 3.6 kΩ, R_2 = 2 kΩ, R_b = 3.6 kΩ, R_z = 20 kΩ, V_z = 6.55 V, R_L = 1 kΩ, V_i = 20 V to 30 V, $V_{i\text{-nominal}}$ = 25 V, $\beta_1 = \beta_2$ = 100, one can apply the inequality constraint given above. Thus,

$$(\beta_1+1)\frac{1}{R_b}+(\beta_1+1)\beta_2\frac{1}{R_1}+\frac{1}{R_1}=2.834$$

$$\frac{1}{R_L}=\frac{1}{1000}=1\times10^{-3}$$

It can be observed that $1/R_L$ is much less than 2.834 and hence the output voltage can be considered to be reasonably independent of R_L. However if R_L approaches the short circuit condition, V_o will lose regulation.

Second-Level Modeling

Till now the insights about the series regulator were obtained based on the idealized first-level models for the transistors and the Zener diode. However, there are many non-idealities in the semiconductor devices. If the dominant non-idealities are included in the device models, then the insights obtained and the conclusions about the circuit will be closer to reality. Figure 4.23 depicts the second-level models for the transistors and the Zener diode. For the NPN transistor shown in Figure 4.23(a), the second-level equivalent circuit is shown in Figure 4.23(b). The base-spreading resistance and the base-to-emitter junction potential are included in the equivalent circuit. For the Zener diode of Figure 4.23(c), the second-level equivalent circuit is shown in Figure 4.23(d). The Zener resistance, r_z in the breakdown region is also included in the equivalent circuit.

In the series regulator circuit of Figure 4.18, the second-level equivalent circuits are used to replace the transistors and the Zener diode. The equivalent circuit of the series regulator with the second-level models introduced is shown in Figure 4.24. Like in the case of the first-level modeling, Thevenin's theorem is applied to the base circuit of transistor Q_2. The reduced circuit after applying Thevenin's theorem is shown in Figure 4.25.

Referring to Figure 4.25, the base current of Q_2 is given as

$$i_{b2}=\frac{V_o[R_2/(R_1+R_2)]-V_{be2}-V_z-r_z[(V_i-V_z)/(R_z+r_z)]}{r_{b2}+R_1/\!/R_2+(\beta_2+1)r_z} \tag{4.42}$$

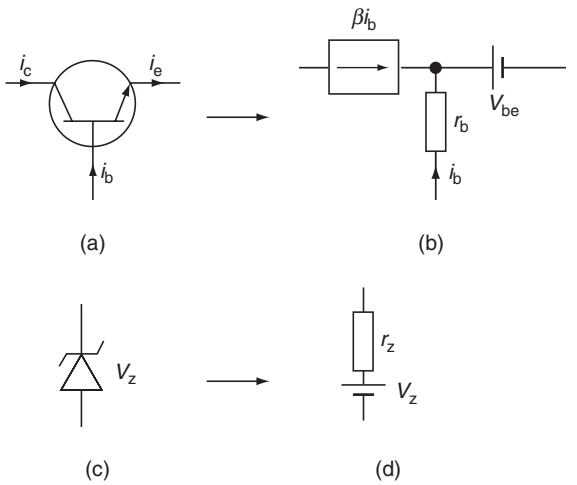


Figure 4.23 | (a) NPN transistor; (b) NPN transistor second-level model; (c) Zener diode; (d) second-level model of Zener diode.

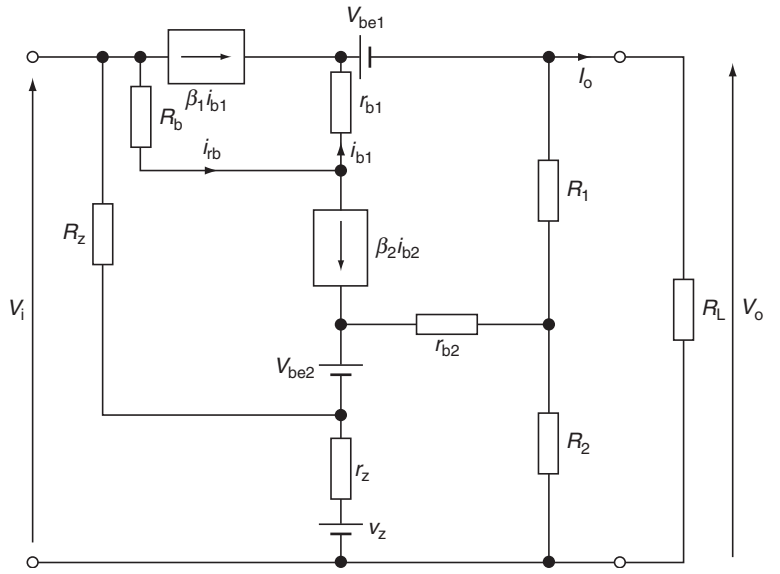


Figure 4.24 | Equivalent circuit of series regulator with second-level models.

The current through R_b is given as

$$i_{rb} = \frac{V_i - (V_o + V_{be1} + i_{b1} r_{b1})}{R_b} \tag{4.43}$$

$$i_{b1} = i_{rb} - \beta_2 i_{b2} \tag{4.44}$$

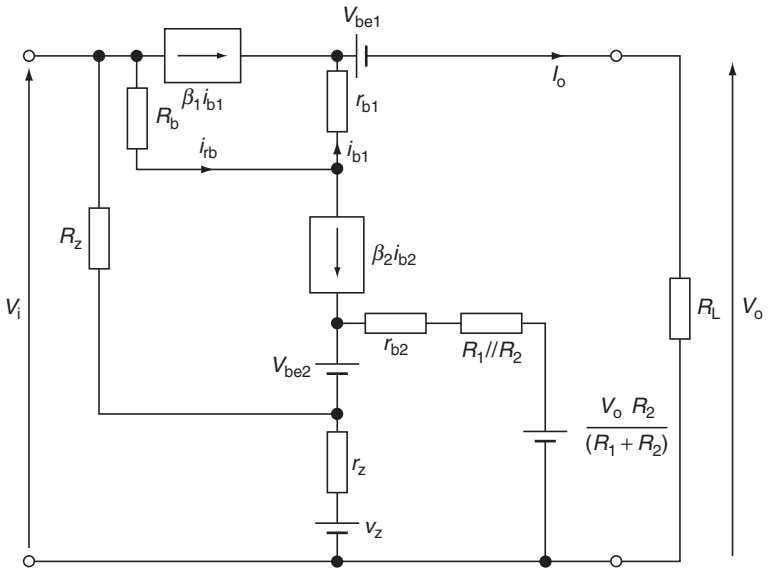


Figure 4.25 | Equivalent circuit of series regulator after circuit reduction.

The output voltage V_o is given as

$$V_o = (i_{b1} + \beta_1 i_{b1}) R_L = (\beta_1 + 1) i_{b1} R_L \tag{4.45}$$

The output relationship can be expressed in the form

$$V_o = k_1 V_i + k_2 V_z + k_3 V_{be1} + k_4 V_{be2} \tag{4.46}$$

where

$$k_1 = \frac{1}{D}\left\{(\beta_1 + 1)\left(\frac{R_L}{R_b + r_{b1}}\right) + \frac{(\beta_1 + 1)\beta_2 R_L [r_z / (R_z + r_z)] R_b}{[r_{b2} + R_1 // R_2 + (\beta_2 + 1) r_z](R_b + r_{b1})}\right\}$$

$$k_2 = \frac{1}{D}\left\{\frac{(\beta_1 + 1)\beta_2 R_L [R_z / (R_z + r_z)] R_b}{[r_{b2} + R_1 // R_2 + (\beta_2 + 1) r_z](R_b + r_{b1})}\right\}$$

$$k_3 = \frac{1}{D}\left\{-\frac{(\beta_1 + 1) R_L}{R_b + r_{b1}}\right\}$$

$$k_4 = \frac{1}{D}\left\{\frac{(\beta_1 + 1)\beta_2 R_L R_b}{[r_{b2} + R_1 // R_2 + (\beta_2 + 1) r_z](R_b + r_{b1})}\right\}$$

and

$$D = \left\{1 + \frac{(\beta_1 + 1) R_L}{R_b + r_{b1}} + \frac{(\beta_1 + 1)\beta_2 R_L [R_2 / (R_1 + R_2)] R_b}{[r_{b2} + R_1 // R_2 + (\beta_2 + 1) r_z](R_b + r_{b1})}\right\}$$

From Eq. (4.46), it is evident that V_o is dependent not only on V_i and V_z, but also on the base–emitter junction voltages of the transistors. It should be noted that when the non-idealities, that is, V_{be1}, V_{be2}, r_z, r_{b1} and r_{b2} are made equal to zero, the output voltage equation reverts back to the equation developed in the first-level model of the previous sub-section.

Influence of V_i on Output Voltage Regulation

From Figure 4.25, it can be observed that there are two parameters that are dependent on V_i and have a strong influence on the output voltage. They are the following:

1. The current i_{rb} which is dependent on V_i, as can be seen from Eq. (4.43).
2. The current through R_z that biases the Zener diode is also dependent on V_i. This current flows through r_z which causes the drop across r_z to change with V_i.

However, in the first-level model, r_z is zero and, therefore, only i_{rb} variations contributed to the output voltage de-regulation.

The solution to this is as follows: The first dependency with respect to i_{rb} can be solved by using a constant-current source circuit as discussed in the first-level model. This is shown in Figure (4.26). It should be noted that even though i_{rb} is a constant and does not vary with variations in V_i, both the Zener voltage V_{z1} and V_{be3} are temperature-dependent. Therefore i_{rb} has a dependency with respect to temperature variations. However, the Zener can be selected to have a positive temperature coefficient and Q_3 base–emitter junction has negative temperature coefficient. If these two devices are appropriately chosen, the temperature dependency of i_{rb} can be significantly reduced.

The second dependency is with respect to the current through r_z. The current through the Zener breakdown resistance r_z is due to contributions from

1. $(\beta_2 + 1)i_{b2}$
2. $\dfrac{V_i - V_z}{R_z + r_z}$

The second contribution, wherein the current flows from the input through R_z to bias the Zener diode, is dependent on input voltage variations. To solve this dependency, the resistor R_z is re-located in such a manner that the Zener bias current is drawn from the output side rather than the input side. As the output is regulated, the Zener bias current becomes independent of the input voltage variations. Figure 4.26 illustrates the re-location of the Zener bias resistor R_z. After this relocation, the current through R_z is given as $(V_o - V_z)/(R_z + r_z)$.

Effect of r_z on Output Regulation

The current through r_z has two components as discussed, viz., $(\beta_2 + 1)i_{b2}$ and the bias current through R_z. The variations in the drop across r_z due to variations in V_i are solved by re-location of R_z. However, the $(\beta_2 + 1)i_{b2}$ component of the current through r_z also varies to de-regulate the output V_o. If this component of current is made constant then the output voltage regulation action is lost. There are only two solutions, viz., (a) make r_z zero which is not practically possible or (b) do not allow the load-dependent component of the current $(\beta_2 + 1)i_{b2}$ to flow through r_z. The latter solution is possible by use of a high impedance amplifier like the op-amp. This solution is depicted in Figure 4.27. The op-amp has a very high input impedance device. The current through the Zener diode is only the bias current component flowing through the resistor R_z and no load-dependent component flows through it. Thus the reference is a constant and independent of load variations.

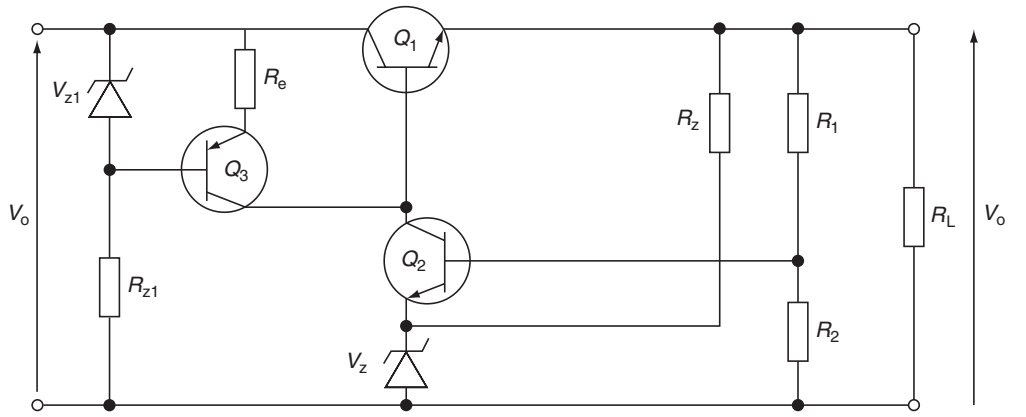


Figure 4.26 Eliminating dependency on input voltage by using constant-current source and relocating R_z.

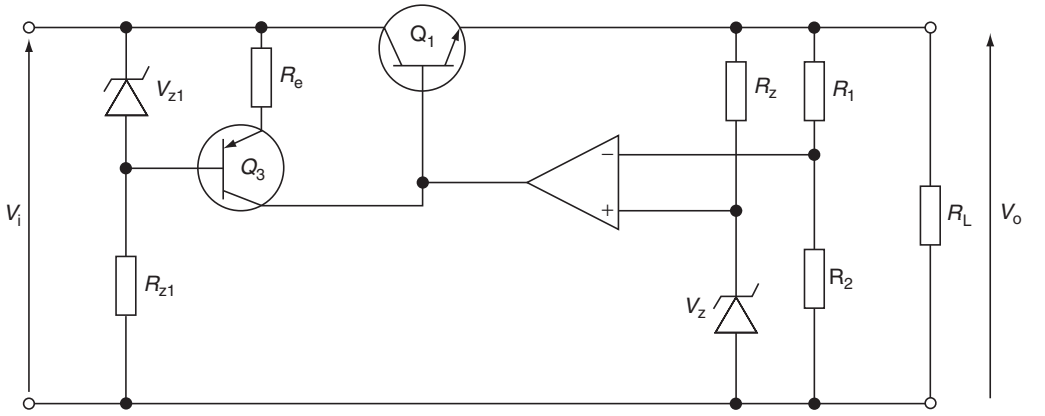


Figure 4.27 Series regulator with op-amp.

There is another advantage in using an op-amp in the series regulator instead of the transistor Q_2. In the first-level model, the range of R_L depends on the inequality

$$\left\{(\beta_1+1)\frac{1}{R_b}+(\beta_1+1)\beta_2\frac{1}{R_1}+\frac{1}{R_1}\right\}\gg\frac{1}{R_L}$$

The gain of the op-amp is very high, of the order of 10^5 and greater. This makes β_2 very large and the above inequality is satisfied for even very low values of R_L. As a consequence the output voltage regulation can be achieved even at very high load, that is, very low R_L. However, in this case the series pass should be rated for high currents.

Influence of Temperature

There are four major parameters that are dependent on temperature. They are:

1. transistor current gains;
2. base–emitter junction potentials of transistors;
3. breakdown voltage of the Zener diode;
4. resistance variations.

From the output voltage in Eq. (4.46), it is evident that all these parameters will significantly affect the output voltage regulation. The transistor current gain β increases with increase in temperature. One way to minimize the effect of variations in β due to change in temperature is to make β as large as possible. The large β will make the output V_o immune to changes in parameters. Another way is to use an op-amp as shown in Figure 4.27. Due to its high gain, the dependency of V_o to changes in parameters due to temperature variations, is significantly reduced.

With regard to the base–emitter junction potential of transistors, the temperature coefficient is negative. The V_{be} of the transistor varies at the rate of −2.2 mV/°C rise in temperature. Zeners below 5.1 V generally have negative temperature coefficient and Zeners above 5.1 V have positive temperature coefficients. Some degree of temperature compensation can be achieved by choosing Zeners with positive temperature coefficient such that they negate the variations of V_{be} with temperature changes.

Another solution to compensate for variations in V_{be} is to use a diode D as shown in Figure 4.28(a). The equivalent circuit for the base drive portion of Q_2 by applying Thevenin's theorem is shown in Figure 4.28(b). The Thevenin voltage for the equivalent circuit of Figure 4.28(b) is given as

$$V_{th} = V_o\left(\frac{R_2}{R_1 + R_2}\right) + V_D\left(\frac{R_1}{R_1 + R_2}\right)$$

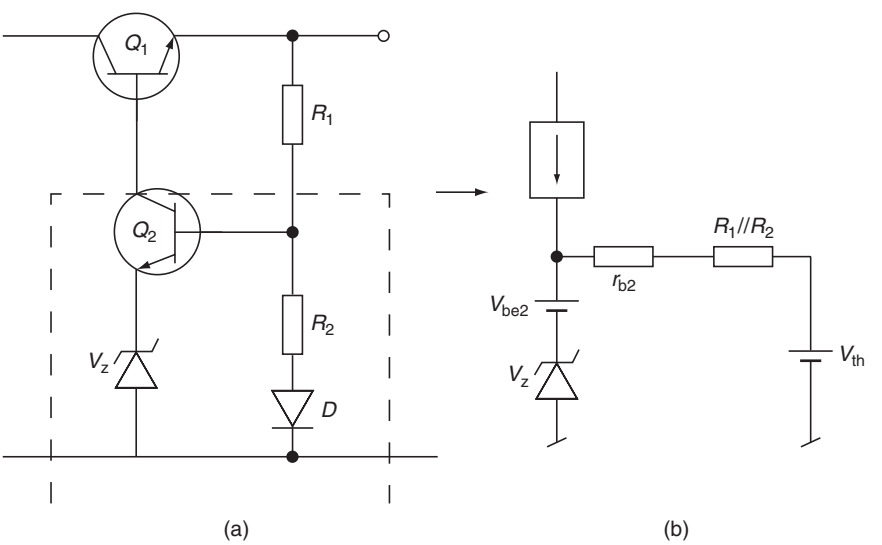


Figure 4.28 | Using diode to compensate for temperature effects.

$$i_{b2} = \frac{V_o[R_2/(R_1+R_2)] + V_D[R_1/(R_1+R_2)] - V_z - V_{be2}}{r_{b2} + R_1 // R_2}$$

From the above equation for i_{b2}, due to opposite signs for the diode voltage and the base–emitter voltage, a certain degree of temperature compensation is achieved between the diode D and V_{be2} of the transistor Q_2.

Alternately, yet another solution is to enclose Q_1, Q_2 and V_z in a constant temperature oven so that there is no drift due to temperature. Though this solution is good for regulation performance, it is at the expense of the efficiency of the series regulator. With regard to resistance variations with temperature, the only solution is to choose very tight tolerance resistors wherein the resistance value variations with temperature are very small.

4.5 Analysis of a Current Regulator

The modeling and analysis methodology followed in the previous section for the series regulator can be applied for any other regulator too. In this section, the analysis is performed for a typical current regulator. In the section "Current Regulation" the circuit and the operational principle of the current regulator has been discussed. A current regulator based on PNP transistor is discussed. This current regulator is also used in the series regulator to improve the output regulation performance as discussed in the previous section. In this section, the analysis is performed on a current regulator with NPN transistor whose circuit schematic is as shown in Figure 4.29.

The operation of this current regulator is exactly similar to that described in the section "Current Regulation". The voltage across R_e is maintained constant and independent of the input voltage V_i. The constant voltage across R_e ensures a constant current through R_e. This current flows as the output current I_o through R_L. In a manner similar to that discussed for the series regulator, the first-level equivalent circuit of the current regulator can be obtained by applying the first-level models for the transistor and the Zener diode as illustrated in Figure 4.19. The first-level model for the current regulator is shown in Figure 4.30.

From the model of Figure 4.30, the voltage across R_e is V_z. The current through R_e is given as

$$I_{Re} = \frac{V_z}{R_e} = (\beta + 1) i_b \tag{4.47}$$

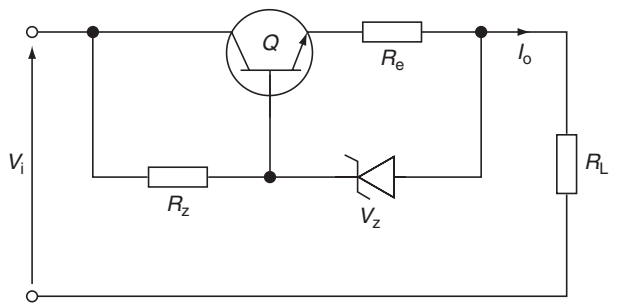


Figure 4.29 | Current regulator circuit.

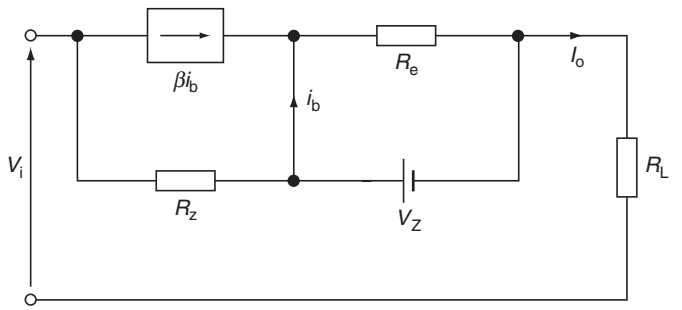


Figure 4.30 | Equivalent circuit of the current regulator with idealized models.

From Eq. (4.47), it is evident that a constant-current source is established as V_z is a constant. Referring again to Figure 4.30, it can be observed that

$$V_o = I_o R_L \tag{4.48}$$

$$i_{Rz} = \frac{V_i - V_z - V_o}{R_z} \tag{4.49}$$

$$i_b = \frac{I_o}{\beta + 1} \tag{4.50}$$

$$i_z = i_{Rz} - i_b \tag{4.51}$$

The output current I_o is given as

$$I_o = \frac{V_z}{R_e} + i_z \tag{4.52}$$

Substituting Eqs. (4.48)–(4.51) into Eq. (4.52), one obtains

$$I_o = \frac{(V_z / R_e) - [I_{Re} / (\beta + 1)] + [(V_i - V_z) / R_z]}{1 / (R_L / R_z)} \tag{4.53}$$

The output current I_o can be re-arranged as

$$I_o = k_1 V_i + k_2 V_z \tag{4.54}$$

where

$$k_1 = \frac{1 / R_z}{1 + (R_L / R_z)}$$

$$k_2 = \frac{(1 / R_e) - [1 / R_e(\beta + 1)] - (1 / R_z)}{1 + (R_L / R_z)}$$

It can be observed from the output current relationship of Eq. (4.54) that the output current I_o is dependent on V_i and the load resistance R_L. It should be noted that R_z should be much larger in value than the resistor R_e to get effective current regulation. This can be achieved by using higher current gain transistors such that the base side currents become very small.

Line Regulation

The term line regulation for current regulators is associated with the variations of the output current I_o for variations in the input voltage V_i. The line-regulation coefficient is expressed as

$$\frac{\partial I_o}{\partial V_i} = \frac{1/R_z}{1+(R_L/R_z)} = \frac{1}{R_L + R_z} \tag{4.55}$$

The variation of output current for variations in the input voltage can be significantly reduced by increasing the value of R_z. This implies that high β transistors should be selected such that the base side currents are small, thereby allowing choice of higher values for R_z.

Range of R_L

In the case of the current regulators, R_L can easily extend up to short circuit at the lower end. However, at the higher end, if R_L is infinite, that is, open circuited, then no I_o flows. Therefore there is an upper limit on the value of the load resistor up to which the current regulation is effective. The load regulation is ensured as long as the following inequality is satisfied.

$$R_z \gg R_L \tag{4.56}$$

Problem 4.3

Consider a current regulator wherein V_i = 20 V to 30 V, $V_{i\text{-nominal}}$ = 25 V, V_z = 5.1 V, R_e = 10 Ω, R_L = 10 Ω, R_z = 3.6 kΩ and β = 100. How much should be the output voltage for regulation to be effective?

Solution

From Eq. (4.54),

$$I_o \approx 0.5 \text{ A}$$

The line-regulation coefficient is given as

$$\frac{\partial I_o}{\partial V_i} = \frac{1}{R_L + R_z} = 2.77 \times 10^{-4}$$

If R_L increases to say 100 Ω, then I_o cannot be maintained at 0.5 A. If I_o were to be 0.5 A, then $V_o = I_o R_L$ = 50 V. This is not possible as the maximum input voltage itself is 30 V. Therefore the output current will invariably lose regulation. For regulation to be effective, the output voltage should be less than the input voltage.

Second-Level Modeling

The second-level models that include the dominant non-idealities of the semiconductor devices can be introduced to obtain the second-level model of the current regulator circuit. The second-level models of the transistor and the Zener diode are as discussed in the previous section and are illustrated in Figure 4.23. Substituting these models in the current regulator circuit, the second-level model equivalent circuit is obtained as shown in Figure 4.31.

Here the dominant transistor non-idealities like the base-spreading resistance and the base–emitter junction potential are introduced in the transistor model and the Zener breakdown resistance is introduced in the Zener model.

Referring to the equivalent circuit of Figure 4.31, the current through R_e is given as

$$I_{Re} = \frac{V_z + i_z r_z - i_b r_b - V_{be}}{R_e} \tag{4.57}$$

The output current I_o is given as

$$I_o = I_{Re} + i_z = \frac{V_z + i_z r_z - i_b r_b - V_{be}}{R_e} + i_z \tag{4.58}$$

From the output current given by Eq. (4.58), it can be noted that the term $i_z r_z$ causes loss of regulation due to the dependence of i_z on V_i. The Zener current is given as

$$i_z = i_{Rz} - i_b$$

Here i_b is constant for a constant I_o. However, i_{Rz} depends on V_i and is given as

$$i_{Rz} = \frac{V_i - V_z - i_z r_z - V_o}{R_z} \tag{4.59}$$

If i_{Rz} is obtained from a constant-current source, then it becomes independent of V_i. This will make i_z independent of V_i which in turn will make the drop $i_z r_z$ independent of V_i. As a consequence the output

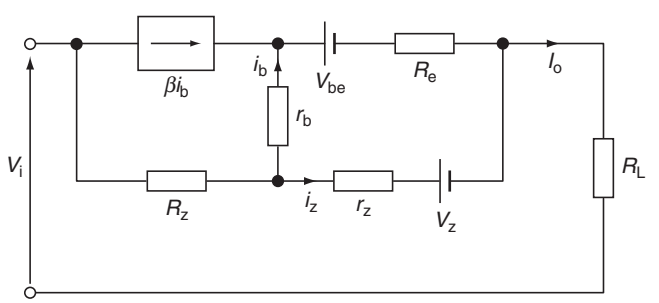


Figure 4.31 | Equivalent circuit of the current regulator with dominant device non-idealities.

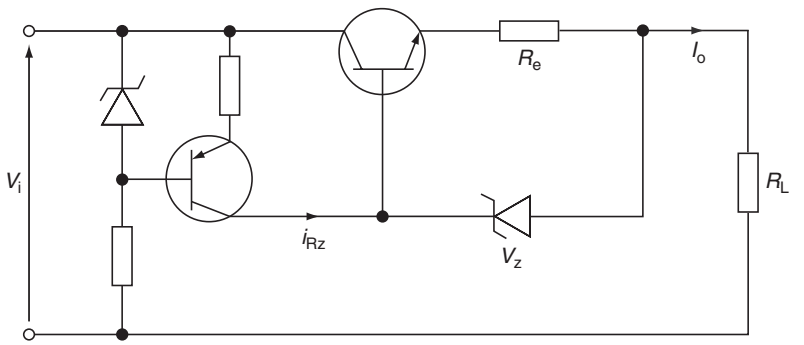


Figure 4.32 Current regulator circuit with R_z replaced by a constant-current source.

current I_o becomes independent of V_i. The current regulator with the R_z replaced by a constant-current source is shown in Figure 4.32.

The effect of temperature on V_{be}, V_z and β is similar to the discussion in the previous section on series regulator. The solutions suggested are equally applicable here too.

CONCLUDING REMARKS

The linear regulators are a class of DC–DC converters that are dissipative, have low efficiency, are bulky due to large heat sink requirements and also have low power density. However, not withstanding all the above disadvantages and more, when it comes down to the quality of the output voltage, there is no switched-mode converter topology that can even closely match the linear regulator performance.

Thus if an application demands high ripple rejection and low line-, load- and temperature-regulation coefficients, then linear regulator is way beyond any other DC–DC converter and therefore must be the obvious choice. If compactness and high efficiency are the governing constraints for a specific application, the switched-mode converter topologies are the obvious choice.

There are many commercial IC regulators that incorporate all the features discussed in this chapter. Among the commercial IC regulators, the low dropout regulators are popular for applications demanding high performance and efficiency. The input–output differential voltage drop across the series-pass device is the cause for the loss of efficiency. The load current multiplied by the input–output differential voltage drop across the series-pass device of the regulator is the dominant power dissipation of the regulator. In low dropout regulators this input–output differential voltage is significantly lower around 0.5 V or 0.7 V. As a consequence the efficiency is improved.

LABORATORY EXERCISES

1. Consider the series regulator circuit shown in Figure 4.33. The input is a variable DC power supply that represents the unregulated DC input voltage.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Simulation in SciLAB
- **c.** Hardware bread-boarding

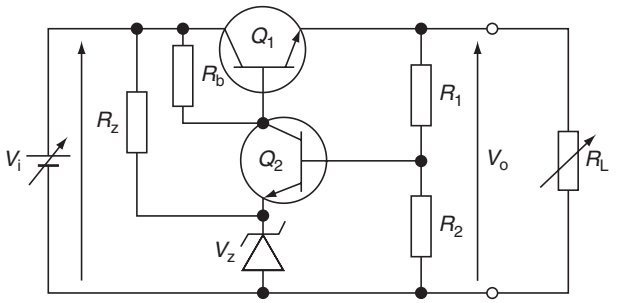


Figure 4.33 | Series regulator circuit.

Tasks for study:

(a) Rig up the series regulator circuit as shown in Figure 4.33 with appropriately selected components values.

(b) Connect an appropriate load resistor across the output terminals.

(c) Keeping the output load resistor fixed, vary the input voltage and tabulate V_i, V_o and I_o.

(d) Keeping the input voltage at a fixed value, vary the load resistor and tabulate V_o and I_o.

(e) From the tabulated results, estimate the line- and load-regulation coefficients.

(f) From the tabulated results, evaluate the efficiency of the linear regulator at various loads and input voltages. Plot the efficiency versus input voltage using load as a parameter.

(g) Vary R_1/R_2 ratio and observe the effect on the output voltage.

(h) Measure the voltage across R_b for different values of the input voltage. Plot the current through R_b versus V_i.

(i) Instead of connecting one end of R_b to the input voltage, connect R_b to another fixed voltage source. Vary the input voltage source keeping load fixed and tabulate V_i, V_o and I_o. From the tabulated results, estimate the line-regulation coefficient. How does it compare with the line-regulation estimate of Step (e)?

(j) Instead of connecting one end of R_z to the input voltage, connect R_z to the output. Evaluate line and load regulation and compare with those evaluated in Step (e).

(k) Incorporate a current protection circuit with the help of an additional NPN transistor and a current sense resistor.

(l) Replace Q_2 with a Darlington BJT (like TIP122). Measure the effect on the range of the load resistor.

2. Consider the three-pin IC regulator circuit shown in Figure 4.34. The input is a variable DC power supply that represents the unregulated DC input voltage.

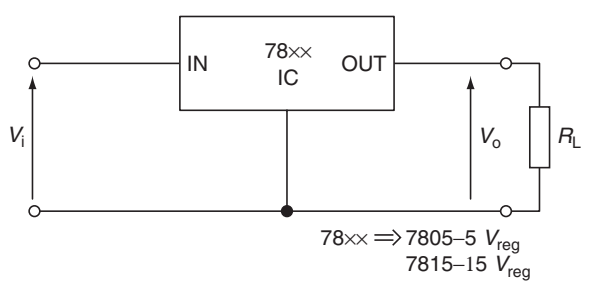


Figure 4.34 | Three-pin 78xx series IC regulator.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware bread-boarding

Tasks for study:

(a) Rig up the circuit as shown in Figure 4.34 using either 7805 or 7815 regulator IC.
(b) Study the data sheet and connect an appropriate load resistor across the output terminals.
(c) Keeping the output load resistor fixed, vary the input voltage and tabulate V_i, V_o and I_o.
(d) Keeping the input voltage at a fixed value, vary the load resistor and tabulate V_o and I_o.
(e) From the tabulated results, estimate the line- and load-regulation coefficients.
(f) From the tabulated results, evaluate the efficiency of the linear regulator at various loads and input voltages. Plot the efficiency versus input voltage using load as a parameter.
(g) From the tabulated results plot the efficiency versus input–output differential voltage.
(h) What is the minimum value of the input voltage for which the output is regulated?

3. Consider the 723 IC regulator circuit shown in Figure 4.35. The input is a variable DC power supply (V_{cc}) that represents the unregulated DC input voltage. R_{sc} is the current sense resistor that is used for over-current protection.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware bread-boarding

Tasks for study:

(a) Study the data sheet of 723 regulator IC.
(b) Rig up the IC regulator circuit as shown in Figure 4.35.
(c) Connect an appropriate load resistor across the output terminals (start with 1 kΩ).
(d) Measure the voltage at pin "V_{ref}". What is its value? Does it agree with the data sheet value?
(e) Keeping the output load resistor fixed, vary the input voltage and tabulate V_i, V_o and I_o.

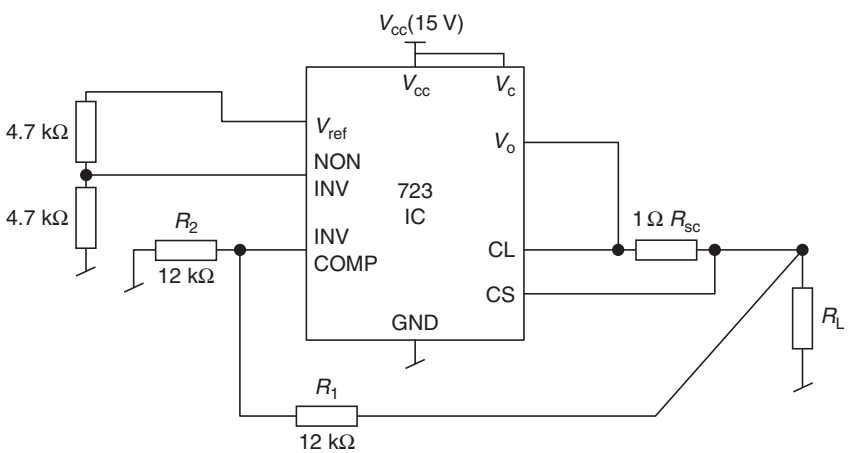


Figure 4.35 | Circuit schematic of 723 IC regulator.

(f) Keeping the input voltage at a fixed value, vary the load resistor and tabulate V_o and I_o.

(g) From the tabulated results, estimate the line- and load-regulation coefficients.

(h) From the tabulated results, evaluate the efficiency of the linear regulator at various loads and input voltages. Plot the efficiency versus input voltage using load as a parameter.

(i) Vary R_1/R_2 ratio and observe the effect on the output voltage.

(j) Change the reference voltage applied to the "NON INV" pin of the IC. What happens to the output?

(k) R_{sc} is the current sense resistor for current protection. Decrease the load resistor value and monitor the voltage across the resistor R_{sc}. At what value of the load resistor does current limit come into operation?

(l) Increase the load current capability by connecting an external BJT. Connect the base of the external BJT to "V_o" pin of the IC, the collector to "V_c" pin of the IC and the emitter to the junction of the "CL" pin and one terminal of R_{sc}. By how much is the load current capability increased?

4. Consider the current regulator circuit shown in Figure 4.36. The input is a variable DC power supply that represents the unregulated DC input voltage.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Simulation in SciLAB
c. Hardware bread-boarding

Tasks for study:

(a) Rig up the current regulator circuit as shown in Figure 4.36 with appropriately selected components values.

(b) Connect an appropriate load resistor across the output terminals.

(c) Keeping the output load resistor fixed, vary the input voltage and tabulate V_i, V_o and I_o.

(d) Keeping the input voltage at a fixed value, vary the load resistor and tabulate V_o and I_o.

(e) From the tabulated results, calculate the variation of the output current with respect to the input voltage variation and load variation.

(f) From the tabulated results, evaluate the efficiency of the linear regulator at various loads and input voltages. Plot the efficiency versus input voltage using load as a parameter.

(g) Change R_e and observe the effect on the output current.

(h) Increase R_o from zero and observe the effect on the output current. At what value of R_o does the output current lose regulation?

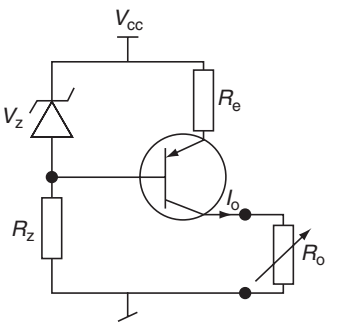


Figure 4.36 | Current regulator circuit.

FILL IN THE BLANKS

1. In linear regulators the power semiconductor devices are operated in the _____ region.
2. The output DC voltage value is always _____ than the input unregulated DC voltage value.
3. Even though the linear regulators have low _____, the quality of the output voltage with respect to the variations in _____, _____ and _____ are better than switched-mode regulator.
4. In a simple shunt regulator, the shunt component is a _____ that is operated in the _____ region.
5. In a shunt regulator, the output voltage is regulated by virtue of the Zener diode being operated in the _____ region.
6. The current absorbs the changes in the _____ and _____ currents.
7. In a shunt regulator, under no-load condition, the whole input current flows into the _____.
8. A transistor or BJT operating in the linear region can be used as a _____ element in a series regulator.
9. If the input voltage is a constant, then the current through the bias resistor R_b of the series regulator can be considered as a _____ current.
10. Current regulators are built using _____ regulators.
11. The no-load condition for the current regulator is when the output resistor value is _____.
12. Efficiency primarily affects the volumetric _____ of the linear regulator.
13. The power dissipated in the series-pass element determines the size of the _____ that needs to be employed.
14. The first-level model for the transistor is a three-wire node with the collector current being a _____ current source of value _____.
15. The first-level model of a Zener diode is a _____ source of value _____.
16. For a series regulator, as the output resistor value approaches infinity, the output voltage is _____ of load.
17. For a series regulator, as the output resistor value approaches zero, the output voltage becomes more and more _____ on load.
18. The second-level model for the transistor includes the _____ and the base–emitter _____.
19. The second-level model of a Zener diode includes the _____.
20. The load regulation of a current regulator is ensured as long as _____.

DESCRIPTIVE QUESTIONS

1. How are voltage scaling and regulation achieved in a linear regulator?
2. What is the efficiency of a generic linear regulator?
3. What are the various linear regulator topologies?
4. Explain the operation of a Zener-based shunt regulator.
5. Discuss the operation of the series regulator.
6. Discuss the effect of input voltage variation on the output voltage.

7. Draw a circuit schematic of a negative voltage regulator and explain its operation.
8. What are the categories of the IC linear regulators? Give examples for each category.
9. What is four-wire connection in linear regulators and why is it needed?
10. Discuss the protection features needed to be incorporated in the linear regulator circuits.
11. Give a current protection circuit for a series linear regulator. If the load over-current limit is to be set at 5 A, then choose the components of the protection circuit.
12. Explain the operation of the crow-bar voltage protection circuit.
13. What are the quantitative measures of a linear regulator that determine its quality?
14. Define line-regulation, load-regulation and temperature-regulation coefficients for a linear regulator.
15. In the analysis of linear regulators, what is first-level modeling?
16. What are the first-level models for BJT and Zener diode?
17. In a series regulator, using the first-level model, what is the relation between output and the base current of the series-pass transistor?
18. In a series regulator, how is the current through the bias resistor R_b made independent of the input voltage?
19. What is the effect of the current flowing through the output sense resistors R_1 and R_2 on the output voltage?
20. What is the range of the load for a series linear regulator?
21. What is the effect of the output resistor R_L on the output voltage regulation?
22. Discuss the constraints on the range of the output resistor, R_L.
23. What are the second-level models for BJT and Zener diode?
24. Using the second-level model of the series regulator, discuss the effect of V_i on the output regulation.
25. Does the Zener breakdown resistance r_z have an effect on the output voltage regulation? Explain.
26. What are the parameters of the series regulator that are affected by temperature? What is the effect of temperature on output voltage regulation?
27. How does one compensate for the temperature variation of the BJT's base–emitter junction potential?
28. How does one compensate for the Zener diode's breakdown voltage variation with temperature?

PROBLEMS

1. A linear regulator is used to regulate an unregulated DC source with nominal voltage of 15 V. The shunt resistor value is 100 Ω and the series resistor value is 250 Ω. A load resistor of 1 kΩ is connected across the output. What is the output voltage? What is the input power drawn from the unregulated input source? What is the efficiency of the linear regulator?
2. For Problem 1, if there is no shunt resistor, then what is the output voltage? What is the input power drawn from the unregulated input source? What is the efficiency of the linear regulator?

3. Consider a Zener shunt regulator circuit wherein the input voltage V_i varies between 10 V and 20 V and the Zener diode used is a 5 V, 1 W device. If the maximum allowable Zener current is 150 mA, then what is the value and power rating of the series resistance used in the regulator?
4. For Problem 3, what is the load resistor range?
5. For the shunt regulator of Problem 3, if the Zener has a breakdown resistance of 0.1 Ω, what is the output voltage variation?
6. Consider the shunt regulator circuit shown in Figure 4.37. The Zener is a 1 W device with the maximum allowable Zener current as 150 mA. The base–emitter voltage drop is 0.6 V. What is the power rating of the series resistance used in the regulator? What is the load resistor range? If the Zener has a breakdown resistance of 0.1 Ω, what is the output voltage variation?
7. For the current regulator circuit shown in Figure 4.38, calculate the value of the load current. What is the upper limit on the output voltage?
8. For Problem 7, what is the range of the load resistor?
9. In a series regulator, the output voltage sensing resistors R_1 and R_2 are 5.6 kΩ each. The output voltage is a regulated 15 V. The Zener

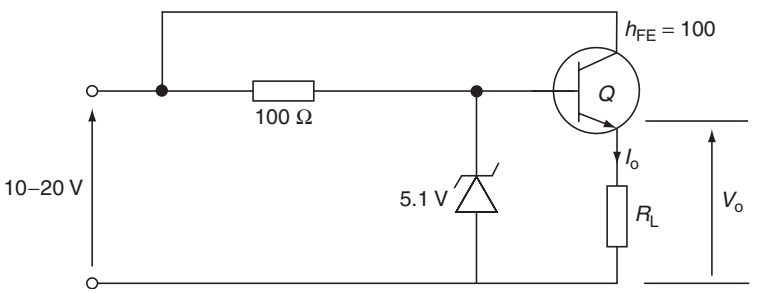


Figure 4.37 | Problem 6.

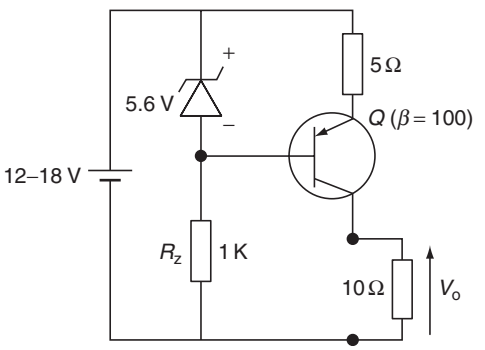


Figure 4.38 | Problem 7.

diode value is 5.1 V. Using the first-level model, calculate the base current through the transistor Q_2.

10. The input voltage of a series regulator varies between 20 V and 30 V. The output is a regulated 15 V DC. If the bias resistor R_b has a value of 2 kΩ, then using the first-level model, estimate the nominal current through the bias resistor.

11. Consider a linear series regulator with $R_1 = 3.6$ kΩ, $R_2 = 2.5$ kΩ, $R_b = 3.6$ kΩ, $R_z = 20$ kΩ, $V_z = 7$ V, $R_L = 1$ kΩ, $V_i = 20$ V to 30 V, $V_{i\text{-nominal}} = 25$ V, $\beta_1 = \beta_2 = 100$. If the output voltage is expressed as $V_o = k_1 V_i + k_2 V_z$, evaluate k_1 and k_2 using the first-level model.

12. In Problem 11, the current gain of Q_1 is increased to 1000. What is the effect on k_1 and k_2? Explain the effect or lack of effect.

13. In Problem 11, the current gain of Q_2 is increased to 1000. What is the effect on k_1 and k_2? Explain.

14. Consider a linear series regulator with $R_1 = 3.6$ kΩ, $R_2 = 2.5$ kΩ, $R_b = 3.6$ kΩ, $R_z = 20$ kΩ, $V_z = 7$ V, $R_L = 1$ kΩ, $V_i = 20$ V to 30 V, $V_{i\text{-nominal}} = 25$ V. Using the first-level model, find the line-regulation coefficient for (a) $\beta_1 = \beta_2 = 100$, (b) $\beta_1 = 1000$ and $\beta_2 = 100$, (c) $\beta_1 = 100$ and $\beta_2 = 1000$.

15. For Problem 14, using the first-level model, find the load-regulation coefficient for (a) $\beta_1 = \beta_2 = 100$, (b) $\beta_1 = 1000$ and $\beta_2 = 100$, (c) $\beta_1 = 100$ and $\beta_2 = 1000$.

16. In the series regulator of Problem 14, if R_b is replaced by a constant-current source regulator, then estimate the load-regulation coefficient for (a) $\beta_1 = \beta_2 = 100$, (b) $\beta_1 = 1000$ and $\beta_2 = 100$, (c) $\beta_1 = 100$ and $\beta_2 = 1000$.

17. In a series regulator, the output voltage sensing resistors R_1 and R_2 are 5.6 kΩ each. The output voltage is a regulated 15 V for a nominal input voltage of 25 V. The Zener diode value is 5.1 V and it is biased with a 20 kΩ resistor from the input. Using the second-level model, calculate the base current through the transistor Q_2. (Assume $V_{be} = 0.6$, $r_b = 0.2$ Ω for BJTs and $r_z = 0.1$ Ω for Zeners. The saturation current gains of Q_1 and Q_2 are 100.)

18. For Problem 17, the input voltage of a series regulator varies between 20 V and 30 V. The output is a regulated 15 V DC. If the bias resistor R_b has a value of 2 kΩ, then using the second-level model, estimate the nominal current through the bias resistor for a load of 1 A.

19. Consider a linear series regulator with $R_1 = 3.6$ kΩ, $R_2 = 2.5$ kΩ, $R_b = 3.6$ kΩ, $R_z = 20$ kΩ, $V_z = 7$ V, $R_L = 1$ kΩ, $V_i = 20$ V to 30 V, $V_{i\text{-nominal}} = 25$ V, $\beta_1 = \beta_2 = 100$. If the output voltage is expressed as $V_o = k_1 V_i + k_2 V_z + k_3 V_{be1} + k_4 V_{be2}$, evaluate k_1, k_2 k_3 and k_4 using the second-level model. (Assume $V_{be} = 0.6$, $r_b = 0.2$ Ω for BJTs and $r_z = 0.1$ Ω for Zeners.)

20. In Problem 19, the current gain of Q_1 is increased to 1000. What is the effect on k_1 and k_2? Explain the effect or lack of effect.

21. In Problem 19, the current gain of Q_2 is increased to 1000. What is the effect on k_1 and k_2? Explain.

22. Consider a current regulator wherein $V_i = 20$ V to 30 V, $V_z = 8.2$ V, $R_e = 10$ Ω, $R_L = 10$ Ω, $R_z = 3.6$ kΩ and $\beta = 100$. What is the constant output current using the first-level model? Apply and compare with the output current estimated with the second-level model.

23. For Problem 22, find the line-regulation coefficient.

ANSWERS

Fill in the Blanks

1. linear
2. lesser
3. efficiency; input voltage; temperature; load
4. Zener diode; reverse breakdown
5. reverse breakdown
6. input; load
7. Zener
8. series-pass
9. constant
10. voltage
11. zero
12. power density
13. heat sink
14. dependent; βi_b
15. DC; V_z
16. independent
17. dependent
18. base-spreading resistance; junction potential
19. breakdown resistance
20. $R_z \gg R_L$

CHAPTER 5

DC–DC Switched-Mode Converters

Learning Objectives

After reading this chapter, you will be able to:

- analyze the steady-state operation of a DC–DC converter.
- design non-isolated and isolated DC–DC converters.
- understand the process of drawing and analyzing waveforms.
- understand soft-switching converters.

In the previous chapter DC–DC converters were discussed wherein the power devices were operated in the linear region. As a consequence the power devices were acting as power dissipators. This results in reduced conversion efficiency. If the power devices are operated either in the cut-off region or in the full-on or saturation region, they act as switches. DC–DC converters that are based on this principle are called the switched-mode converters. The switched-mode DC–DC converters would ideally have 100% efficiency. However, due to switching and conduction losses this value is lower than 100%. After the voltage is switched with a power switch, the voltage has a switched wave shape that needs to be filtered to obtain a DC output voltage. The filtering or averaging is done using non-dissipative components like the inductor and capacitor. Therefore, any switched-mode converter will include at least a power switch that emulates a single pole double throw (SPDT) mechanical switch, an inductor and a capacitor.

5.1 Single Pole Double Throw Switch

A single pole double throw (SPDT) switch is shown in Figure 5.1(a). The switch consists of a pole P and two throws T1 and T2. The pole makes contact with either T1 or T2 at any given instant. Referring to Chapter 1, it is seen that no single power device can behave as an SPDT switch. However, the power devices can behave as a single pole single throw (SPST) switch. Therefore, from the point of view of implementation using the power devices, the SPST switch is available as a single device. The SPDT switch can be constructed using SPST switches as shown in Figure 5.1(b). Here the SPST switch S1 is linked to throw T1 and the SPST switch S2 is linked to throw T2. At any given instant of time it should be noted that either S1 is ON or S2 is ON. Such an operational topology will emulate an SPDT switch.

Few practical realizations of the SPDT switch are shown in Figure 5.2. These realizations are based on the generic topology of Figure 5.1(b). In the circuits of Figure 5.2, the power switch represented as bipolar transistor (BJT) can be replaced by either a metal oxide semiconductor field effect transistor (MOSFET) or

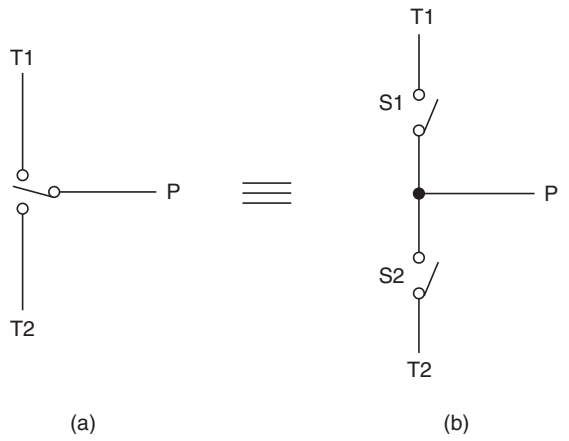


Figure 5.1 | (a) SPDT switch; (b) two SPST switches used to emulate an SPDT switch.

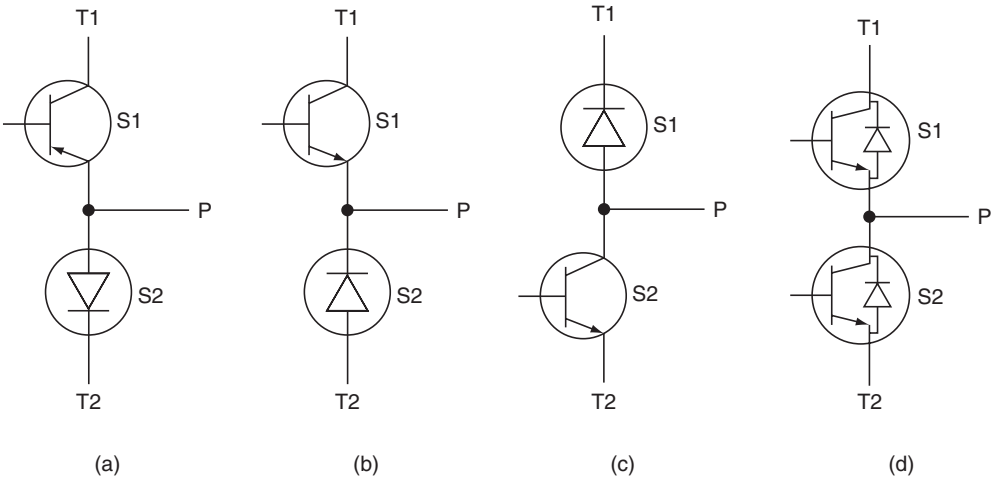


Figure 5.2 | Few realizations of the SPDT switch using power devices.

an insulated gate bipolar transistor (IGBT) without loss of generality. In Figure 5.2(d), the diodes are the internal body diodes of the BJT or MOSFET or IGBT that is used. This is in fact one of the most popular of all topologies and is extensively used in both DC–DC converters and DC–AC inverters.

5.2 The Chopper

The chopper is a power converter wherein the input DC voltage is chopped and sent to the output without incorporating any filters exclusively within the converter. The filtering of the voltage or current is the responsibility of the applied load. The schematic of the chopper is illustrated in Figure 5.3(a). The simplest

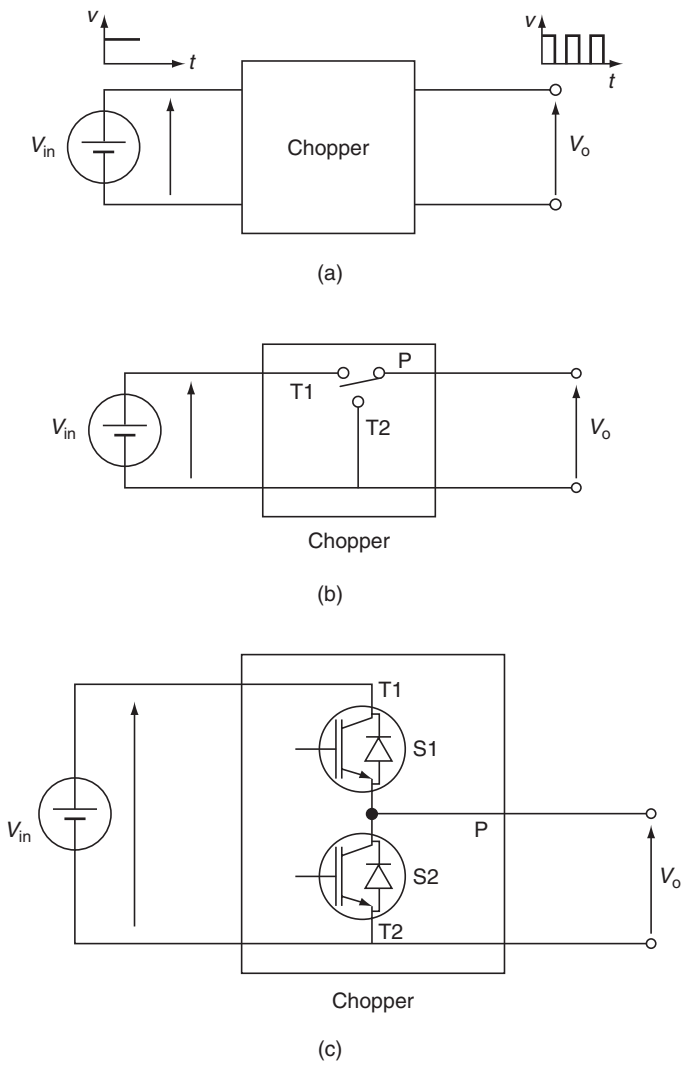


Figure 5.3 Basic chopper: (a) Block schematic; (b) SPDT representation; (c) realization using semiconductor switches.

chopper can be constructed using only one SPDT switch. As the output is an unfiltered chopped waveform, there is no need for an inductor and capacitor. Figure 5.3(b) gives the chopper representation with an SPDT switch and Figure 5.3(c) gives the practical realization of the chopper. The switches S1 and S2 are IGBT switches with internal body diodes. The IGBT switches may be replaced with MOSFET or BJT switches.

The switch S1 is turned ON by appropriate application of the gate-to-emitter pulse. During this time, the switch S2 is in the OFF-state. This duration is for a period T_{on}. For the period T_{off}, the switch S1 is turned OFF and the switch S2 is turned ON. The pole voltage at P is a switching waveform as shown in Figure 5.4.

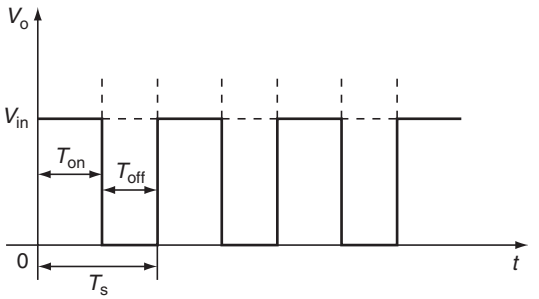


Figure 5.4 | Chopper output waveform.

Referring to Figure 5.4,

$$T_s = T_{on} + T_{off}$$

where T_s is one switching period. A useful and important term that is widely used in the power electronic literature is the duty ratio or the duty cycle. The duty ratio or the duty cycle (D) can be defined as the ratio of the ON-time of switch S1 when the input is connected to the pole to the total switching period. This is given as

$$D = \frac{T_{on}}{T_{on} + T_{off}} = \frac{T_{on}}{T_s} \tag{5.1}$$

The output voltage of the chopper is a switched waveform with an average value of

$$V_o = \frac{1}{T_s}\int_0^{T_s} v_o \mathrm{d}t = \frac{1}{T_s}\int_0^{T_{on}} V_i \mathrm{d}t = \frac{V_i}{T_s} T_{on} = DV_i$$

The chopper circuit is used in applications like DC motor drives wherein the motor's armature inductance filters the armature current flowing through it. Figure 5.5(a) shows the chopper connected to the armature circuit of a separately excited DC motor. In Figure 5.5(b), the DC motor is replaced by an equivalent circuit that acts as the load on the chopper. The DC motor is represented by the armature resistance R_a, the armature inductance L_a and the speed-dependent back emf source e_b.

The DC motor can operate in any one of the four operating zones as shown in the v–i characteristic of Figure 5.6, namely, (a) forward motoring mode or first quadrant operation, (b) reverse motoring mode or the third quadrant operation, (c) forward generation mode or second quadrant operation, (d) reverse generation mode or fourth quadrant operation. The chopper drive shown in Figure 5.5 can work in only first and second quadrants as it cannot supply negative average output voltage. Therefore, for this chopper drive the motor can operate in only two modes: (a) forward motoring and (b) forward generation. These modes are indicated in Figures 5.7 and 5.8.

In the forward motoring mode the power flows from the DC source V_i to the motor shaft. The armature voltage and current are positive wherein the armature current i_a flows into the motor +ve terminals. In this mode the back emf source e_b acts as a sink or a load. In one chopper switching period T_s, S1 is ON during

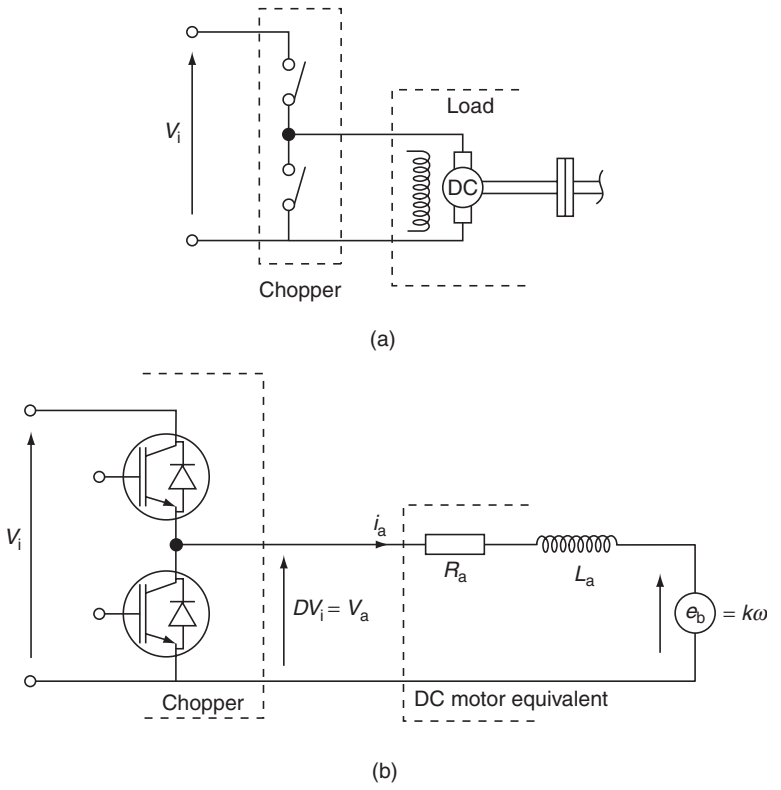


Figure 5.5 (a) Chopper connected to a separately excited DC motor; (b) circuit schematic with electrical equivalent of DC motor.

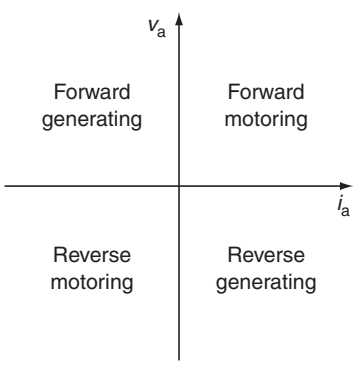


Figure 5.6 Four quadrant operation of DC motor.

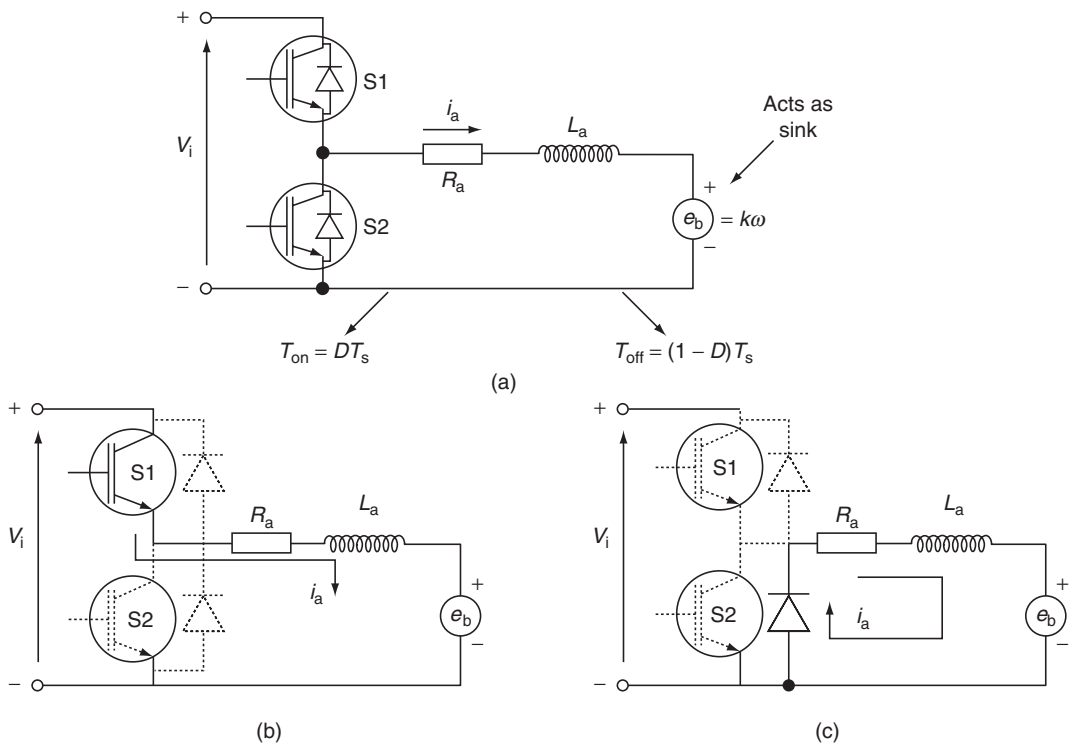


Figure 5.7 Forward motoring: (a) Forward motoring mode; (b) operative circuit during DT_s; (c) operative circuit during $(1 - D)T_s$.

DT_s period and S2 is ON during $(1 - D)T_s$ period. For DT_s period, the operative portion of the circuit is shown in Figure 5.7(b). As the armature current i_a flows into the motor terminals as shown, the IGBT of switch S1 is operative and is indicated in Figure 5.7(b). For $(1 - D)T_s$ period, the operative portion of the circuit is shown in Figure 5.7(c). During this time, only the diode of S2 is operative wherein the armature current freewheels through the diode as shown in Figure 5.7(c).

In the forward-generation mode the power flows from the motor shaft to the DC source V_i. This occurs during braking of the motor wherein the mechanical inertia translates to the electric domain and charges up the DC source V_i. The armature voltage is positive as in forward motoring mode, but the armature current is negative in a direction wherein the armature current i_a flows out of the motor +ve terminals. In this mode the back emf source e_b acts as a generator. In one chopper switching period T_s, the switch S1 is ON during DT_s period and the switch S2 is ON during $(1 - D)T_s$ period. For $(1 - D)T_s$ period, the operative portion of the circuit is shown in Figure 5.8(c). During this time, the IGBT of S2 is operative wherein the armature current flows as shown in Figure 5.8(c) and charges up the energy in the armature inductance. For DT_s period, the operative portion of the circuit is shown in Figure 5.8(b) wherein the energy of the armature inductance discharges through the diode and into the DC supply V_i as indicated in Figure 5.8(b).

A single basic chopper consisting of the SPDT switch can handle only two quadrants of operation wherein the DC applied voltage being constant, the chopper output current (i_a) in the DC motor example

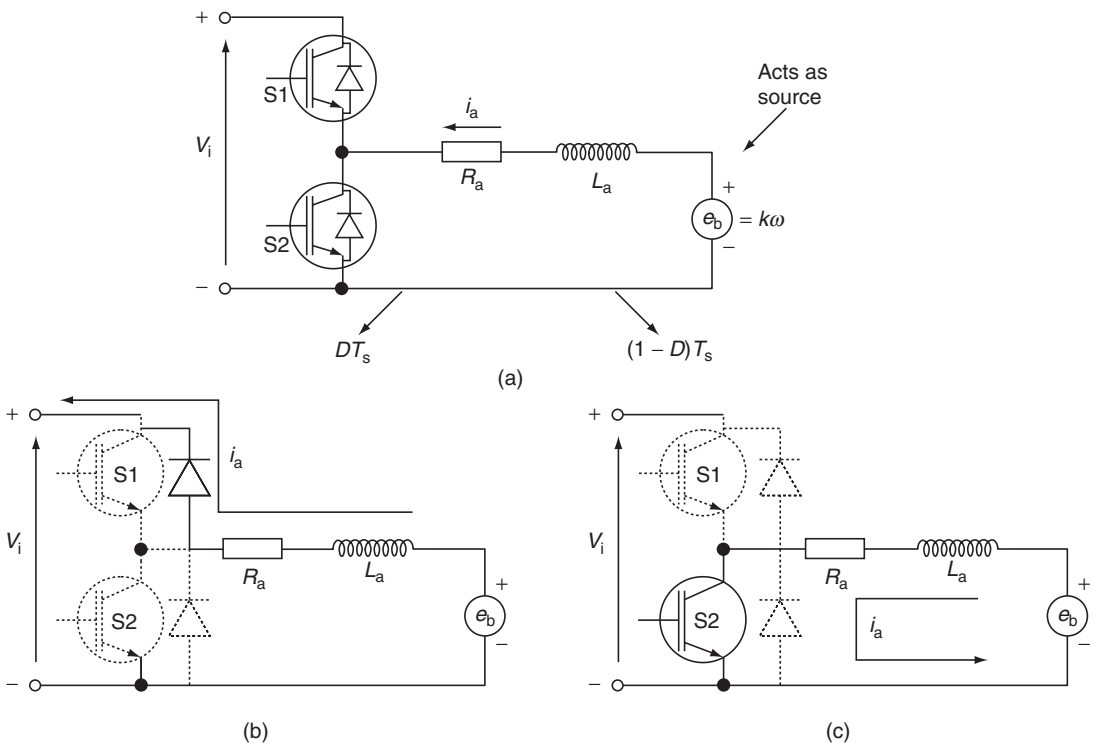


Figure 5.8 | Forward regeneration: (a) Forward-regeneration mode; (b) operative circuit during DT_s; (c) operative circuit during $(1 - D)T_s$.

illustrated above can flow in either direction. However, if all four quadrant operation is needed wherein both the chopper output voltage and current are bi-directional, then two basic choppers are required. The voltage across the poles of the two SPDT switches, V_{P1P2} can handle bi-directional voltages as well. The schematic of the dual–four quadrant chopper is shown in Figure 5.9. Figure 5.10 shows the application of the dual chopper to the DC motor load for four quadrant operation.

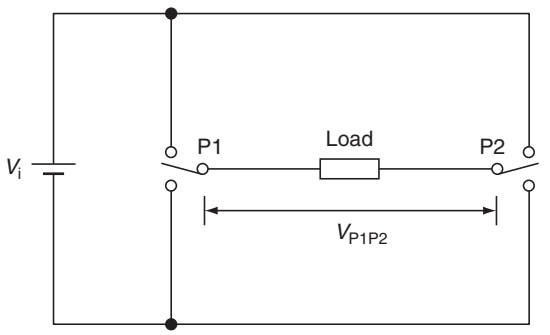


Figure 5.9 | Schematic of the dual chopper using two SPDT switches.

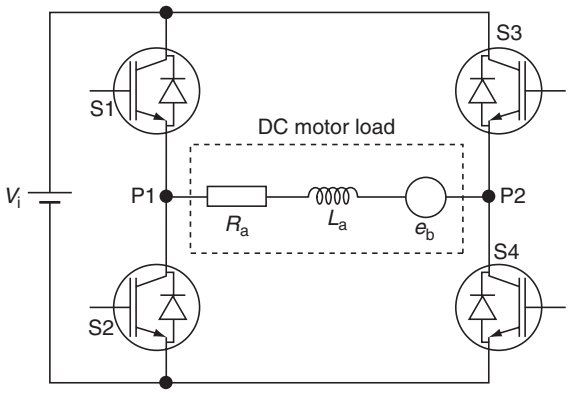


Figure 5.10 Circuit schematic of the dual chopper with DC motor load across the poles P1 and P2.

5.3 DC Steady-State Principles

The basic chopper is combined with the inductors and the capacitors to obtain the DC output. The following sections will discuss DC–DC converters using chopper, inductors and capacitors in various topological combinations. The converters will be discussed with respect to their steady-state operation. The dynamic modeling of the converters will be dealt with in Chapter 8.

The analyses presented in this chapter are based on the following two fundamental physical laws that govern the steady-state operation of the switched-mode power converters:

1. **Volt-Second Balance:** The inductor in a circuit will not support an average DC voltage across it. This implies that under steady-state conditions,

$$\int_0^{T_s} v_L \, dt = 0$$

where v_L is the voltage across the inductor and T_s is the switching period. This means that the product of voltage and time or, in other words, the area under the inductor voltage curve in one period should be zero under equilibrium conditions. This is called the *volt-second balance* for inductors or windings.

2. **Amp-Second Balance or Charge Balance:** The capacitor in a circuit will not allow an average DC current to pass through it. This implies that under steady-state conditions,

$$\int_0^{T_s} i_C \, dt = 0$$

where i_C is the current through the capacitor and T_s is the switching period. This means that the product of the current through the capacitor and time or, in other words, the area under the capacitor current curve in one period should be zero under equilibrium conditions. This is called the *amp-second balance* for capacitors. As amp-second is the same as the charge on the capacitor, it is also called *charge balance*. This implies that there is neither decrease nor increase in the charge on the capacitor under equilibrium conditions.

5.4 The Three Basic Converters

This section discusses three converters synthesized using topological variations of a basic chopper, an inductor and a capacitor. While synthesizing any converter the following two rules must be strictly followed to obtain a practically realizable DC–DC converter:

1. The inductor stores energy by virtue of the current flowing through it $[(1/2)Li^2]$. This implies that the current through the inductor cannot change instantaneously. Therefore, the inductor should be placed such that the inductor current does not become discontinuous during any instant or mode of operation.
2. The capacitor stores energy by virtue of the voltage across it $[(1/2)Cv^2]$. This implies that the voltage across the capacitor cannot change instantaneously. Therefore, the capacitor should be placed such that the capacitor voltage does not become discontinuous during any instant or mode of operation.

Based on the application of the rules mentioned above, three converter variants are discussed in this section. The three variants are (a) step-down or the buck converter, (b) step-up or the boost converter and (c) step-up/down or buck–boost converter. These three converters are called the primary converters as other converters can be synthesized using these converters and so they form the basic building blocks.

Step-Down or Buck Converter

Topology

The step-down or the buck converter as the name suggests, gives an output voltage that is always smaller than the input voltage. The buck converter topology is illustrated in Figure 5.11. Figure 5.11(a) shows the buck converter consisting of an SPDT switch, an inductor that is connected to the pole of the SPDT switch and a capacitor that is connected across the output as shown. The input DC voltage source V_i is connected across the throws, T1 and T2, of the SPDT switches. Observe that the inductor is connected to the pole of the SPDT switch where there is no discontinuity in the current. Figure 5.11(b) shows the buck converter with the SPDT switch replaced by two SPST switches S1 and S2. Figure 5.11(c) shows the operative circuit during the period DT_s of a switching period. During this time, S1 is ON and S2 is OFF. The pole P is connected to the input V_i. The inductor current flows from the input source to the output as shown. Figure 5.11(d) shows the operative circuit during the period $(1 - D)T_s$ of the switching period. During this time S1 is OFF and S2 is ON. The inductor current cannot change instantaneously. The inductor current starts to decrease. This negative di/dt of the inductor current develops sufficient voltage $(L\,di/dt)$ with a polarity such as to drive the inductor current through the switch S2. This is called the freewheeling action of the inductor current.

Operation

The operation of the buck converter may be analyzed by visualizing the waveforms of the various signals of the circuit. In general for any converter, the following signals may be considered important for analysis:

1. voltage across the inductor;
2. current through the inductor;
3. load current;
4. current through the capacitor;
5. input source current;
6. voltages across switches S1 and S2;
7. currents through switches S1 and S2.

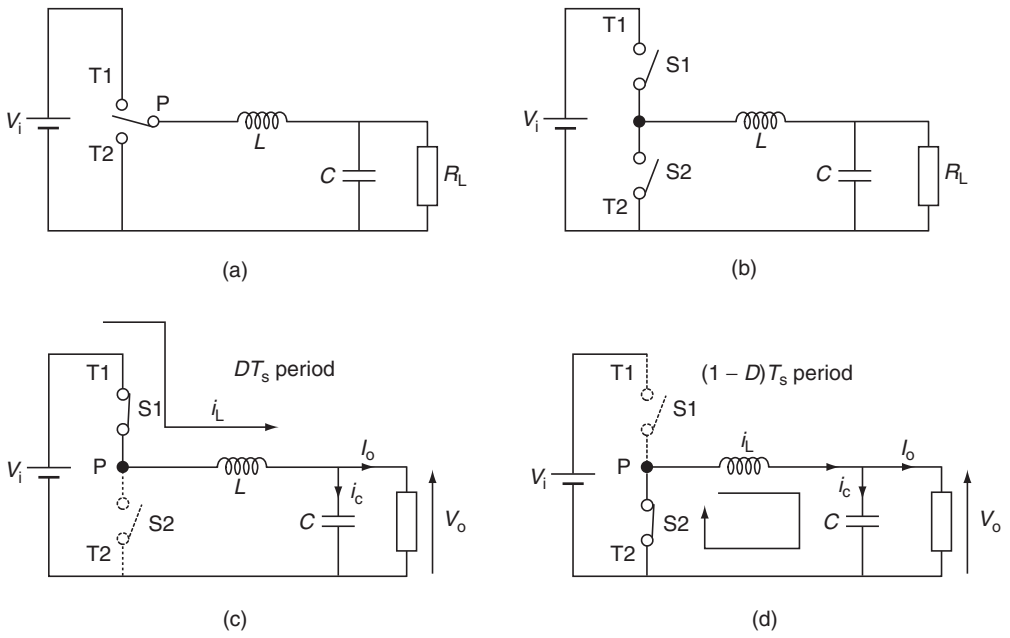


Figure 5.11 Step-down or buck converter: (a) Buck converter with a SPDT switch, an L and a C; (b) SPDT switch is replaced with two switches S1 and S2; (c) operative circuit during the period DT_s; (d) operative circuit during the period $(1 - D)T_s$.

Figure 5.12 shows the plot of the various signals for the buck converter under steady-state operating conditions.

DT_s Duration The switch S1 is ON and S2 is OFF during the period DT_s. The voltage v_L across the inductor during this time is $V_i - V_o$. Both V_i and V_o are DC quantities and hence $V_i - V_o$ will be a DC quantity during the period DT_s. From the Faraday's law, the inductor voltage and current are related as, $v_L = L(\mathrm{d}i_L / \mathrm{d}t)$. This implies that the current through the inductor is given as

$$i_L = \frac{1}{L}\int v_L \mathrm{d}t = \frac{1}{L}\int (V_i - V_o)\mathrm{d}t = \frac{V_i - V_o}{L}\int \mathrm{d}t$$

As $V_i - V_o$ is a constant, the inductor current i_L rises linearly with a slope of $(V_i - V_o)/L$ as indicated in Figure 5.12. During this time the current through S1 is the same as the inductor current. The voltage across S2, which is OFF, is V_i as the pole P is connected to the input source V_i.

$(1 - D)T_s$ Duration During the time when S1 is OFF and S2 is ON, the voltage v_L across the inductor during this time is $-V_o$. Again as V_o is a DC quantity, the inductor current i_L falls linearly with a slope of $-V_o/L$ as indicated in Figure 5.12. During this time, the current through S2 is the same as the current through the inductor. The voltage across S1, which is OFF, is V_i as the pole P is connected to the circuit ground. The inductor current is composed of two parts: (a) the DC component (I_o) and (b) the AC component (i_c). The DC component cannot flow through the capacitor, therefore the DC component flows into the load R_L. The AC component alone will flow through the capacitor. Therefore, the capacitor current i_c is the inductor current minus the load DC component.

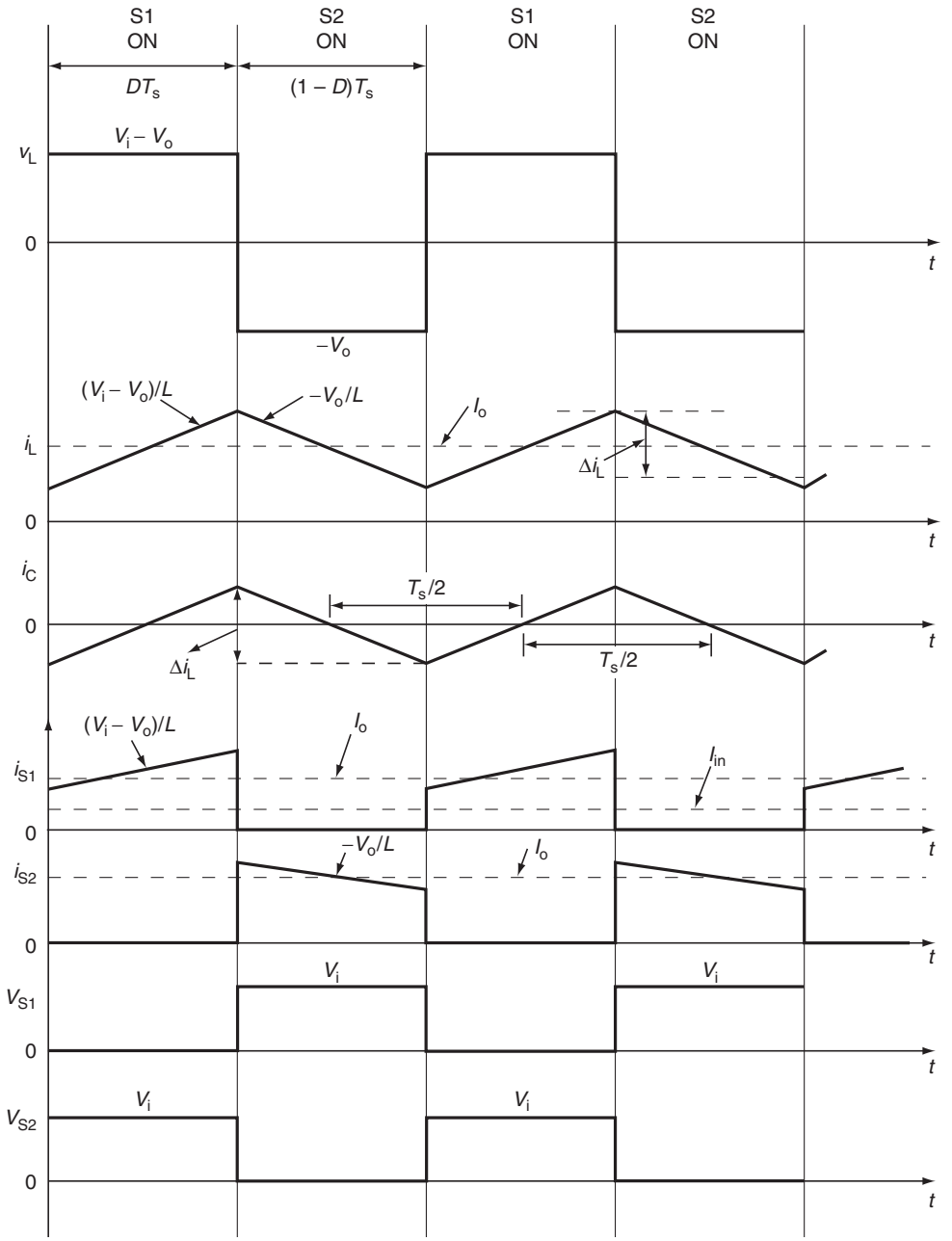


Figure 5.12 | Waveforms of various signals of a buck converter.

Semiconductor Switches

The buck converter schematic of Figure 5.11 shows only the conceptual SPDT and SPST switches. The SPST switches may be replaced with power semiconductor switches as shown in Figure 5.13. Referring to the operating modes given in Figures 5.11(c) and (d) and the OFF-state voltages across S1 and S2 given in Figure 5.12, the switch S1 can be either a BJT or a MOSFET or an IGBT and the switch S2 is a diode that is required for freewheeling the inductor current when S1 is OFF. This implementation is shown in Figure 5.13(a). However, the power semiconductor switches like MOSFETs and IGBTs are mostly available along with their body diodes in-built; therefore, in order to reduce the inventory, the switch configuration of Figure 5.13(b) is also common. Here the IGBT part of S1 and the diode part of S2 are the operative components for this topology.

Steady-State Analysis

Based on the application of the volt-second balance and charge balance rules, the input–output relationship under steady-state conditions can be obtained.

With Ideality Constraints The analysis is performed by assuming the following:

1. The switches S1 and S2 have no ON-state drops.
2. The winding resistance of the inductor is zero.
3. The equivalent series resistance (ESR) of the capacitor is zero.
4. There are no losses in the converter implying that the efficiency is 100%.

Input–output relationship To obtain the input–output voltage relationship, apply the volt-second balance rule to the inductor. This implies that the area under the inductor voltage curve in one period under steady-state conditions should be zero. Referring to Figure 5.12,

$$(V_i - V_o)DT_s + (-V_o)(1-D)T_s = 0$$

Simplifying the above equation, one obtains

$$V_o = DV_i \tag{5.2}$$

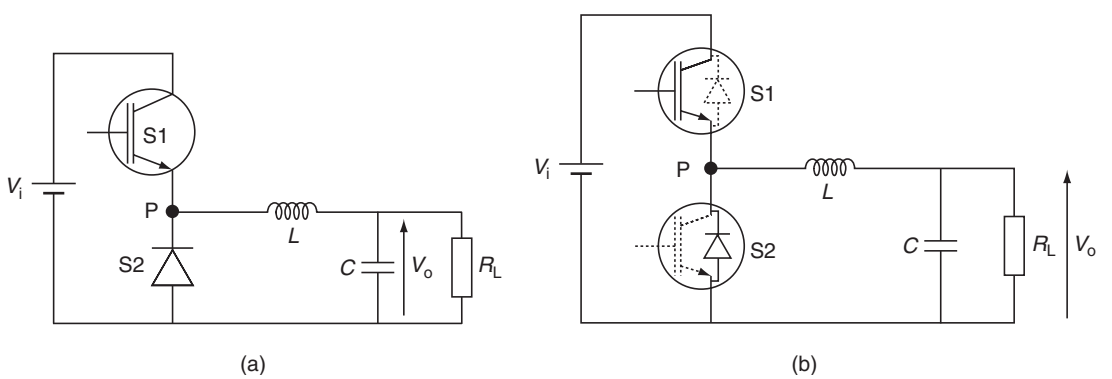


Figure 5.13 | Realization of buck converter with power semiconductor switches.

The duty ratio (D) as defined in Eq. (5.1) can take on values between the closed interval 0 and 1. This means that the output voltage V_o is always less than the input voltage V_i. *Hence this topology is called a step-down converter or buck converter*. To obtain the input–output current relationship one can use the 100% efficiency constraint. This implies that

$$V_i I_{in} = V_o I_o \tag{5.3}$$

Using Eq. (5.2) in Eq. (5.3), one obtains

$$I_{in} = DI_o \tag{5.4}$$

Inductor value The value of the inductor L is calculated based on the amount of current ripple Δi_L that the designer would like to allow for a given application. Referring to the waveforms given in Figure 5.12, the slope of the inductor current waveform during the period DT_s is given as

$$\frac{di_L}{dt} = \frac{\Delta i_L}{\Delta t} = \frac{\Delta i_L}{DT_s} = \frac{V_i - V_o}{L} \tag{5.5}$$

Likewise the slope of the inductor current waveform during the period $(1 - D)T_s$ is given as

$$\frac{di_L}{dt} = \frac{\Delta i_L}{\Delta t} = \frac{\Delta i_L}{(1-D)T_s} = \frac{-V_o}{L} \tag{5.6}$$

The voltages V_i and V_o are known from the converter specifications. D can be obtained from the input–output relationship of Eq. (5.2). The switching frequency f_s is a design choice. From this $T_s = 1/f_s$ is obtained. Δi_L is also a design choice. Though this is chosen based on the application, the common choice in most cases is $\Delta i_L = 10\%$ of I_o. The only unknown L can be calculated from either Eq. (5.5) or Eq. (5.6). Between these two equations the latter is preferred as it is independent of V_i which is an unregulated voltage. Taking the absolute value of the current slope from Eq. (5.6), the value of inductor is given as

$$L = \frac{V_o(1-D)}{\Delta i_L f_s} \tag{5.7}$$

To regulate V_o with variations in V_i, the value of D should be changed. Thus if V_{imax} is the maximum input voltage swing, then D_{min} will be the corresponding minimum duty cycle to obtain a specified V_o. If V_{imin} is the minimum input voltage swing, then D_{max} will be the corresponding maximum duty cycle to obtain the specified V_o. Thus, for a regulated V_o,

$$V_o = V_{imax} \times D_{min} = V_{imin} \times D_{max} \tag{5.8}$$

The values V_o, V_{imax} and V_{imin} are known from specifications of the converter. D_{max} is a design choice that is dependent on the specific application. Then

$$D_{min} = \frac{V_o}{V_{imax}} = \frac{V_{imin} \times D_{max}}{V_{imax}}$$

The inductor current should have a current ripple that is less than or equal to the specified Δi_L. Therefore the inductor value should be calculated at the duty ratio corresponding to D_{min}. Thus,

$$L = \frac{V_o(1 - D_{min})}{\Delta i_L f_s} \tag{5.9}$$

Capacitor value The capacitor value is calculated by applying the amp-second rule. Referring to the capacitor current waveform given in Figure 5.12, if there is to be no charge build up or charge reduction in the capacitor, then the area under the capacitor current curve in one period should be zero. The area under the positive portion of the current curve implies charging of the capacitor and that under the negative portion implies discharging of the capacitor. For charge balance both these areas should be equal. The change is the capacitor charge ΔQ is given by the area under either the positive portion or the negative portion of the capacitor current curve. Thus,

$$\Delta Q = C\Delta V_o = \frac{1}{2} \times \text{Base} \times \text{Height} = \frac{1}{2}\frac{T_s}{2}\frac{\Delta i_L}{2} = \frac{\Delta i_L}{8 f_s} \tag{5.10}$$

$$C = \frac{\Delta i_L}{8\Delta V_o f_s} \tag{5.11}$$

Here ΔV_o is known from the output ripple specification. Δi_L and f_s are design choices as mentioned earlier. The value of C should be greater than the value calculated by Eq. (5.11).

Effect of Non-Idealities Till now, the analysis assumed the switches, inductor and capacitor as ideal. However, practical components are far from ideal. The switches have ON-state drops, the inductor has winding resistance, the capacitor has the ESR that contributes to output voltage ripple and reduction in efficiency, inductor core material has core losses that reduce the efficiency and so on. As a consequence, the output of the practical buck converter will not exactly agree with the input–output relationship derived with the ideality constraints. Figure 5.14 shows the buck converter schematic with few non-idealities included. The ON-state drops of the switches are included as voltage sources in series with the SPDT switches. The source V_{cesat} in series with the IGBT switch represents the ON-state drop of S1 when S1 is ON. The source V_D in series with the diode switch represents the ON-state drop of S2 when S2 is ON. The resistance r_L represents the inductor winding resistance and r_{esr} represents the capacitor ESR.

To derive a more exact input–output voltage relationship, one will have to apply the volt-second and amp-second balance to the circuit of Figure 5.14 wherein the non-idealities are incorporated in the circuit schematic. The design of the capacitor will also get affected due to the presence of the non-idealities especially the ESR non-ideality. The output voltage ripple will contain contributions from the voltage variation due to ESR and that due to the change of charge as discussed earlier. Therefore for a specified output voltage ripple, the voltage ripple contribution due to change of charge should be much lesser than when ESR is absent. This implies that the capacitance value will be higher for a given output voltage ripple.

Step-Up or Boost Converter

Topology

The step-up or the boost converter, as the name suggests, gives an output voltage that is always higher than the input voltage. The boost converter topology is illustrated in Figure 5.15. Figure 5.15(a) shows the boost converter consisting of an SPDT switch, an inductor that is connected to the pole of the SPDT

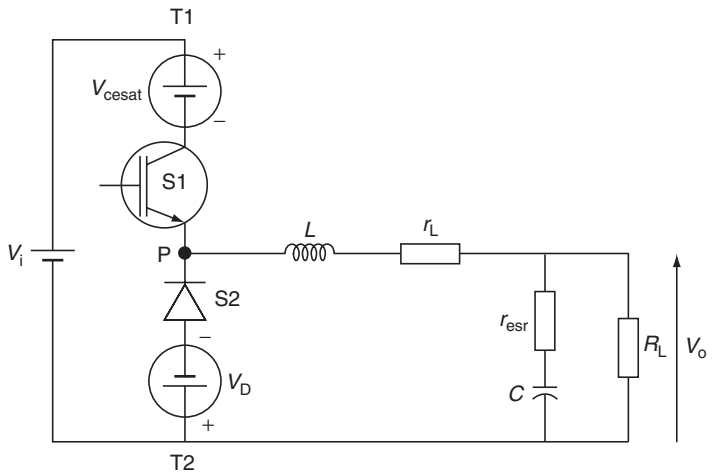


Figure 5.14 | Buck converter with non-idealities included.

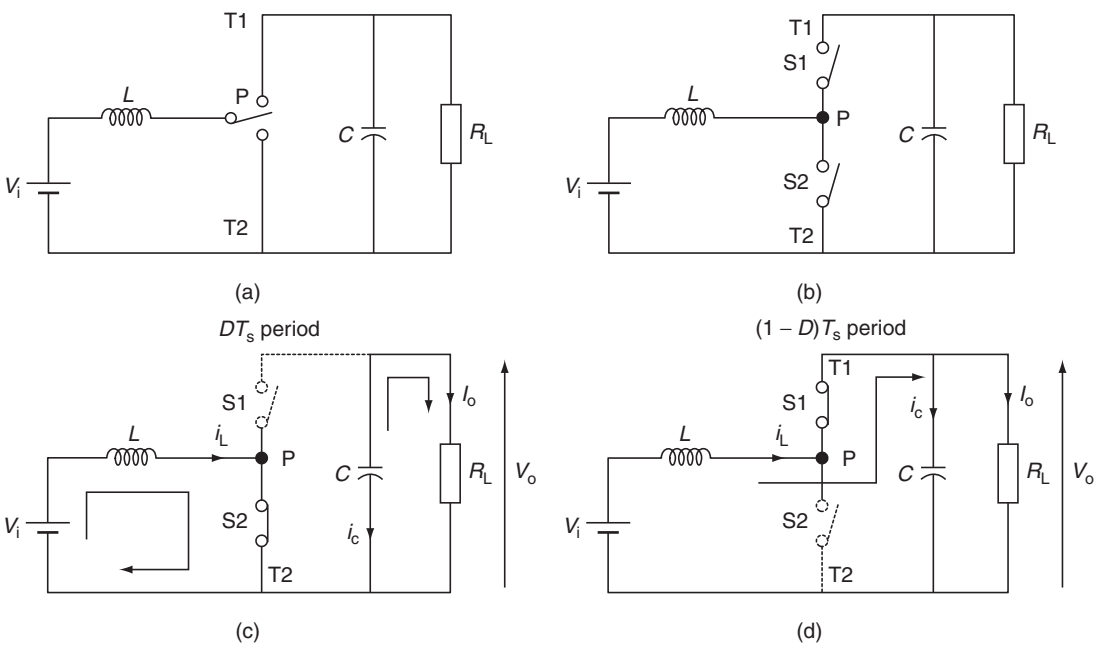


Figure 5.15 | Step-up or boost converter: (a) Boost converter with a SPDT switch, an L and a C; (b) SPDT switch is replaced with two switches S1 and S2; (c) operative circuit during the period DT_s; (d) operative circuit during the period $(1 - D)T_s$.

switch and a capacitor that is connected across the throws of the SPDT switch as shown. Observe that here also the inductor is connected to the pole of the SPDT switch where there is no discontinuity in the current. However note that the inductor is connected between the pole and the input source in the boost configuration whereas in the buck configuration the inductor is connected between the pole and the output. Figure 5.15(b) shows the boost converter with the SPDT switch replaced by two SPST switches S1 and S2. Figure 5.15(c) shows the operative circuit during the period DT_s of a switching period. During this time, S2 is ON and S1 is OFF. The pole P is connected to the circuit ground. The inductor current flows from the input source to energize the inductor. Figure 5.15(d) shows the operative circuit during the period $(1 - D)T_s$ of the switching period. During this time S1 is ON and S2 is OFF. The inductor current cannot change instantaneously. The inductor current starts to decrease. This negative di/dt of the inductor current develops sufficient voltage ($L di/dt$) with a polarity such as to drive the inductor current through S1 to charge the output capacitor.

Operation

The operation of the boost converter is analyzed by visualizing the waveforms of the various signals of the circuit. The following signals are considered important for analysis:

1. voltage across the inductor;
2. current through the inductor;
3. load current;
4. current through the capacitor;
5. input source current;
6. voltages across switches S1 and S2;
7. currents through switches S1 and S2.

Figure 5.16 shows the plot of the various signals of the boost converter under steady-state operating conditions.

DT_s Duration The switch S1 is OFF and S2 is ON during the period DT_s. The voltage v_L across the inductor during this time is V_i which is a DC quantity. From the Faraday's law, the inductor voltage and current are related as $v_L = L(di_L / dt)$. This implies that the current through the inductor is given as

$$i_L = \frac{1}{L}\int v_L \, dt = \frac{1}{L}\int V_i \, dt = \frac{V_i}{L}\int dt$$

As V_i is a constant, the inductor current i_L rises linearly with a slope of V_i/L as indicated in Figure 5.16. During this time the current through S2 is the same as the inductor current. The voltage across S1, which is OFF, is V_o as the pole P is connected to the circuit ground.

$(1 - D)T_s$ Duration During this time when S1 is ON and S2 is OFF, the voltage v_L across the inductor is $V_i - V_o$. In the boost converter, V_o is always greater than V_i and therefore $V_i - V_o$ is a negative quantity. The inductor current i_L falls linearly with a slope of $(V_i - V_o)/L$ as indicated in Figure 5.16. During this time, the current through S1 is the same as the current through the inductor. The voltage across S2, which is OFF, is V_o as the pole P is connected to the output V_o.

The inductor current is composed of two parts: (a) the DC component (I_{in}) and (b) the AC component (Δi_L). Here also the inductor current ripple is a design choice. The S1 switch current (i_{S1}) also has two distinct components as shown in Figure 5.16: (a) the average or DC component (I_o) and (b) the AC component

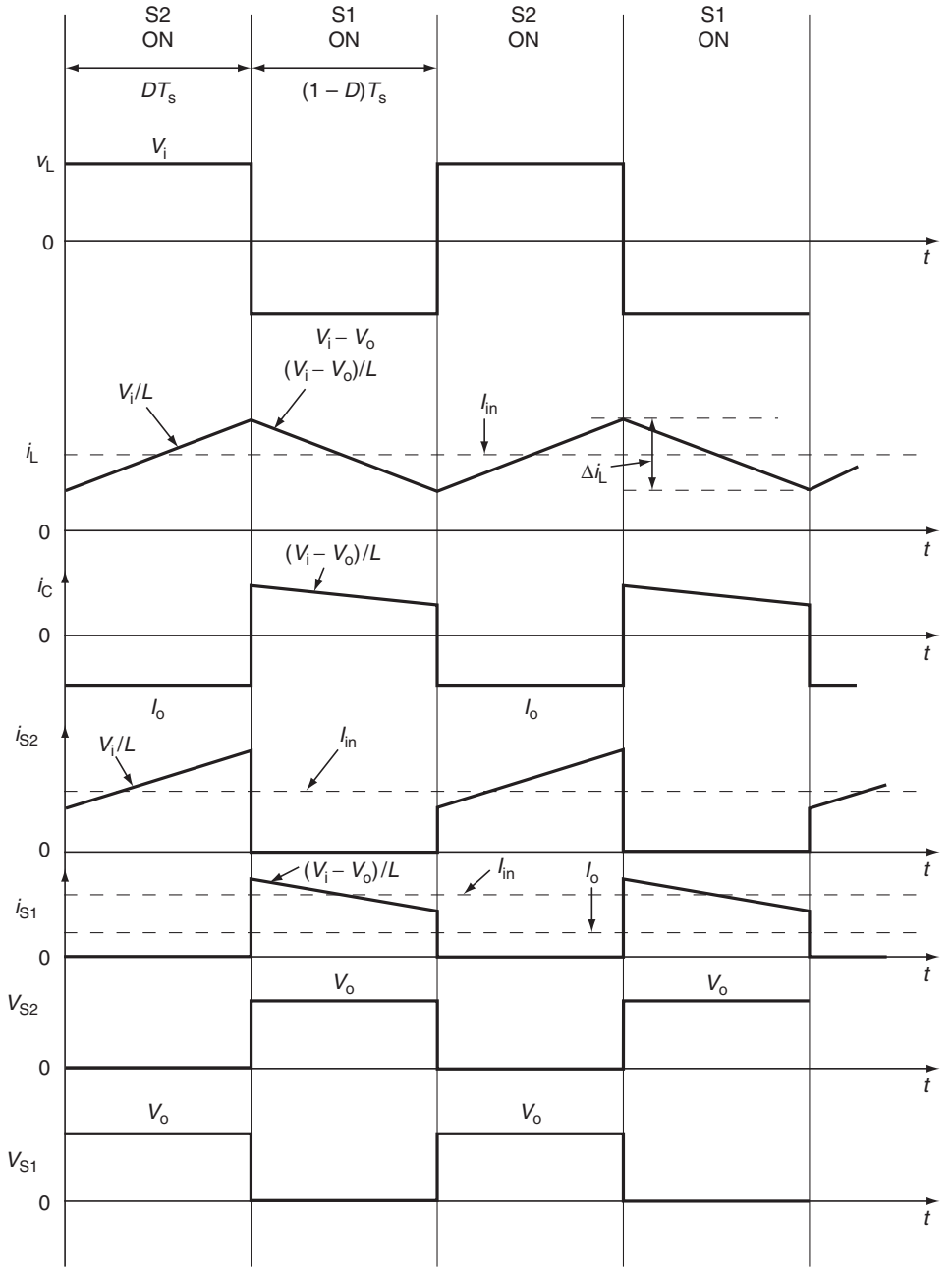


Figure 5.16 | Waveforms of various signals of a boost converter.

(i_c). The DC component cannot flow through the capacitor, therefore it flows into the load R_L. The AC component alone will flow through the capacitor. Therefore, the capacitor current i_c is the S1 switch current minus the load DC component.

Semiconductor Switches

The boost converter schematic of Figure 5.15 shows only the conceptual SPDT and SPST switches. The SPST switches may be replaced with power semiconductor switches as illustrated in Figure 5.17. Referring to the operating modes given in Figures 5.15(c) and (d) and the OFF-state voltages across S1 and S2 given in Figure 5.16, the switch S2 can be either a BJT or a MOSFET or an IGBT and the switch S1 is a diode that turns ON when S2 is switched OFF by virtue of the inductor voltage (Ldi/dt) build up. This implementation is shown in Figure 5.17(a). However, the power semiconductor switches like MOSFETs and IGBTs are mostly available along with their body diodes in-built; therefore, the switch configuration of Figure 5.17(b) is also common. Here the IGBT of S2 and the diode of S1 are the operative components of each switch in this topology.

Steady-State Analysis

Based on the application of the volt-second balance and charge balance rules, the input–output relationship under steady-state conditions can be obtained.

With Ideality Constraints The analysis is performed by assuming the following:

1. The switches S1 and S2 have no ON-state drops.
2. The winding resistance of the inductor is zero.
3. The ESR of the capacitor is zero.
4. There are no losses in the converter implying that the efficiency is 100%.

Input–output relationship To obtain the input–output voltage relationship, apply the volt-second balance rule to the inductor. This implies that the area under the inductor voltage curve in one period under steady-state conditions should be zero. Referring to Figure 5.16,

$$(V_i)DT_s + (V_i - V_o)(1-D)T_s = 0$$

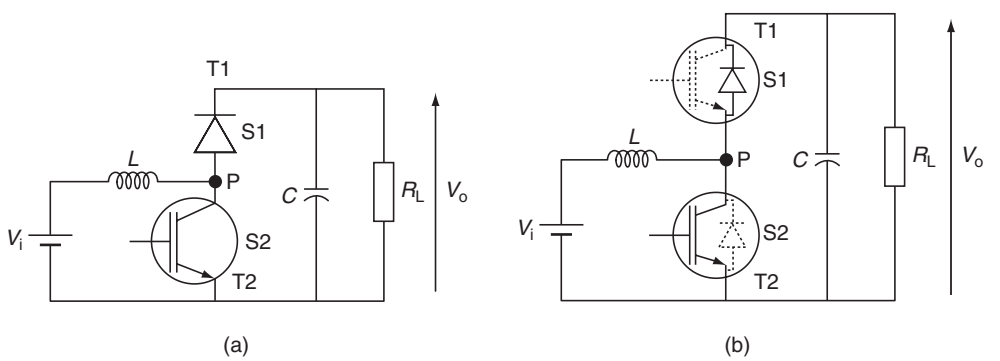


Figure 5.17 | Realization of boost converter with power semiconductor switches.

Simplifying the above equation, one obtains

$$V_o = \frac{V_i}{1-D} \tag{5.12}$$

The duty ratio D, as defined in Eq. (5.1), can take on values between the closed interval 0 and 1. This means that the output voltage V_o is always greater than the input voltage V_i. *Hence this topology is called a step-up converter or boost converter*. Note that V_o is infinite when D is unity. This however is not true. It should be noted that the above input–output relationship is derived considering steady-state operating conditions. This means that the switches S1 and S2 are switching and have reached an equilibrium condition in terms of the ON and OFF times. If D is unity then S2 is always ON and S1 is always OFF. This further implies that on the input side, the inductor current will rise till the inductor saturates and the input current will be limited only by the winding resistance and the ON-state impedance of S2. The output capacitor is disconnected from the input side as S1 is OFF. This means that the capacitor will continually discharge into the load R_L till it losses all its charge and V_o becomes zero. Thus when D is unity, the output voltage V_o will in actuality become zero as Eq. (5.12) is no longer valid. To obtain the input–output current relationship one can use the 100% efficiency constraint. This implies that

$$V_i I_{in} = V_o I_o \tag{5.13}$$

Using Eq. (5.12) in Eq. (5.13), one obtains

$$I_{in} = \frac{I_o}{1-D} \tag{5.14}$$

Inductor value The value of the inductor L is calculated on the basis of the amount of current ripple Δi_L that the designer would like to allow for a given application. Referring to the waveforms given in Figure 5.16, the slope of the inductor current waveform during the period DT_s is given as

$$\frac{di_L}{dt} = \frac{\Delta i_L}{\Delta t} = \frac{\Delta i_L}{DT_s} = \frac{V_i}{L} \tag{5.15}$$

Likewise the slope of the inductor current waveform during the period $(1 - D)T_s$ is given as

$$\frac{di_L}{dt} = \frac{\Delta i_L}{\Delta t} = \frac{\Delta i_L}{(1-D)T_s} = \frac{V_i - V_o}{L} \tag{5.16}$$

Here V_i and V_o are known from the converter specifications. D can be obtained from the input–output relationship of Eq. (5.12). The switching frequency f_s is a design choice. From this $T_s = 1/f_s$ is obtained. Δi_L is also a design choice. The common choice in most cases is $\Delta i_L = 10\%$ of I_{in}. The output power P_o is known from the converter specifications. From the knowledge of P_o and V_o, I_o is calculated. Using Eq. (5.14), I_{in} can be estimated for a given load. The only unknown L can then be calculated from either Eq. (5.15) or Eq. (5.16). From Eq. (5.15), the value of inductor is given as

$$L = \frac{V_i D}{\Delta i_L f_s} \tag{5.17}$$

To regulate V_o with variations in V_i, the value of D should be varied. Thus if V_{imax} is the maximum input voltage swing, then D_{min} will be the corresponding minimum duty cycle to obtain a specified V_o. If V_{imin} is the minimum input voltage swing, then D_{max} will be the corresponding maximum duty cycle to obtain the specified V_o. Thus, for a regulated V_o,

$$V_o = \frac{V_{imax}}{1 - D_{min}} = \frac{V_{imin}}{1 - D_{max}} \tag{5.18}$$

Here V_o, V_{imax} and V_{imin} are known from specifications of the converter. D_{max} is a design choice that is dependent on the specific application. From Eq.(5.18),

$$D_{min} = 1 - \left(\frac{V_{imax}}{V_{imin}} \right) (1 - D_{max})$$

The inductor value should be calculated with the appropriate duty ratio for a particular input voltage swing. Thus,

$$L = \frac{V_{imax} D_{min}}{\Delta i_L f_s} \tag{5.19}$$

Capacitor value It is calculated by applying the amp-second rule. Referring to the capacitor current waveform given in Figure 5.16, if there is to be no charge build up or charge reduction in the capacitor, the area under the capacitor current curve in one period should be zero. The area under the positive portion of the current curve implies charging of the capacitor and that under the negative portion implies discharging of the capacitor. For charge balance both these areas should be equal. The change is the capacitor charge ΔQ is given by the area under either the positive portion or the negative portion of the capacitor current curve. For the boost converter, geometrically it is easier to find the area under the negative portion of the capacitor current curve. Thus,

$$\Delta Q = C \Delta V_o = I_o D T_s = \frac{I_o D}{f_s} \tag{5.20}$$

$$C = \frac{I_o D}{\Delta V_o f_s} \tag{5.21}$$

ΔV_o is known from the output ripple specification. Δi_L and f_s are design choices as mentioned earlier. The value of C should be greater than the value calculated by Eq. (5.21). Compare this with the capacitor value estimate for the buck converter as given by Eq. (5.11). In the case of the buck converter, the capacitor current is only the inductor ripple current which is only a fraction of the load current whereas in the case of the boost converter, the capacitor has to handle a larger swing in the current through it. Therefore, the capacitor size is bigger in the case of the boost converter as compared to the buck converter.

Effect of Non-Idealities The analysis assumed the switches, inductor and capacitor as ideal. However, practical components are far from ideal. The switches have ON-state drops, the inductor has winding resistance, the capacitor has the ESR that contributes to output voltage ripple and reduction in efficiency. As a consequence, the output of the practical boost converter will not exactly agree with the input–output relationship derived with the ideality constraints. Figure 5.18 shows a boost converter schematic with few non-idealities included. The ON-state drops of the switches are included as voltage sources in series with the SPDT switches. The source V_{cesat} in series with the IGBT switch represents the ON-state drop of S2 when S2 is ON. The source V_D in series with the diode switch represents the ON-state drop of S1 when S1 is ON. The resistance r_L represents the inductor winding resistance and r_{esr} represents the capacitor ESR.

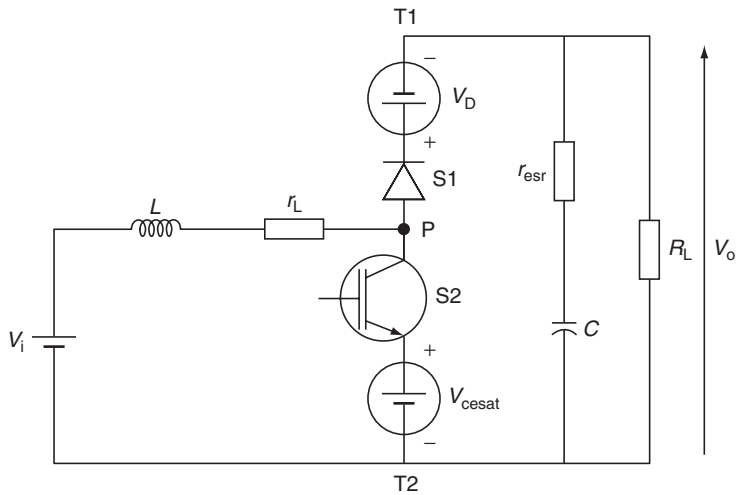


Figure 5.18 | Boost converter with non-idealities included.

To derive a more exact input–output voltage relationship, one will have to apply the volt-second and amp-second balance to the circuit of Figure 5.18 wherein the non-idealities are included in the schematic. The design of the capacitor will also get affected due to the presence of the non-idealities especially the ESR non-ideality. As discussed in the buck converter case, the output voltage ripple will contain contributions from the voltage variation due ESR and that due to the change of charge. Therefore for a specified output voltage ripple, the voltage ripple contribution due to change of charge should be much lesser than when ESR is absent. This would result in a higher value of capacitance than that resulting from Eq. (5.21).

Step-Up/Down or Buck–Boost Converter

Topology

The step-up/down or the buck–boost converter gives an output voltage that can be either higher or lower than the input voltage depending on the operating duty ratio. The buck–boost converter topology is illustrated in Figure 5.19. Figure 5.19(a) shows the buck–boost converter consisting of an SPDT switch, an inductor that is connected to the pole of the SPDT switch and a capacitor that is connected across the output. In all three primary converters, the capacitor is always connected across the output. However, observe that in the case of the buck converter, the inductor is connected between the pole of the SPDT switch and the output; in the case of the boost converter, the inductor is connected between the pole of the SPDT switch and the input; and now in this case of the buck–boost converter, the inductor is connected between the pole of the SPDT switch and the circuit ground. Figure 5.19(b) shows the buck–boost converter with the SPDT switch replaced by two SPST switches S1 and S2. Figure 5.19(c) shows the operative circuit during the period DT_s of a switching period. During this time, S1 is ON and S2 is OFF. The pole P is connected to the input source V_i. The inductor current flows from the input source to energize the inductor. Figure 5.19(d) shows the operative circuit during the period $(1 - D)T_s$ of the switching period. During this time, S1 is OFF and S2 is ON. The inductor current starts to decrease. This negative di/dt of the inductor current develops sufficient voltage (Ldi/dt) with a polarity such as to drive the inductor current through the switch S1 to charge the output capacitor with the polarity as shown. It should be observed that the output voltage polarity is negative with respect to the circuit ground as shown in Figure 5.19.

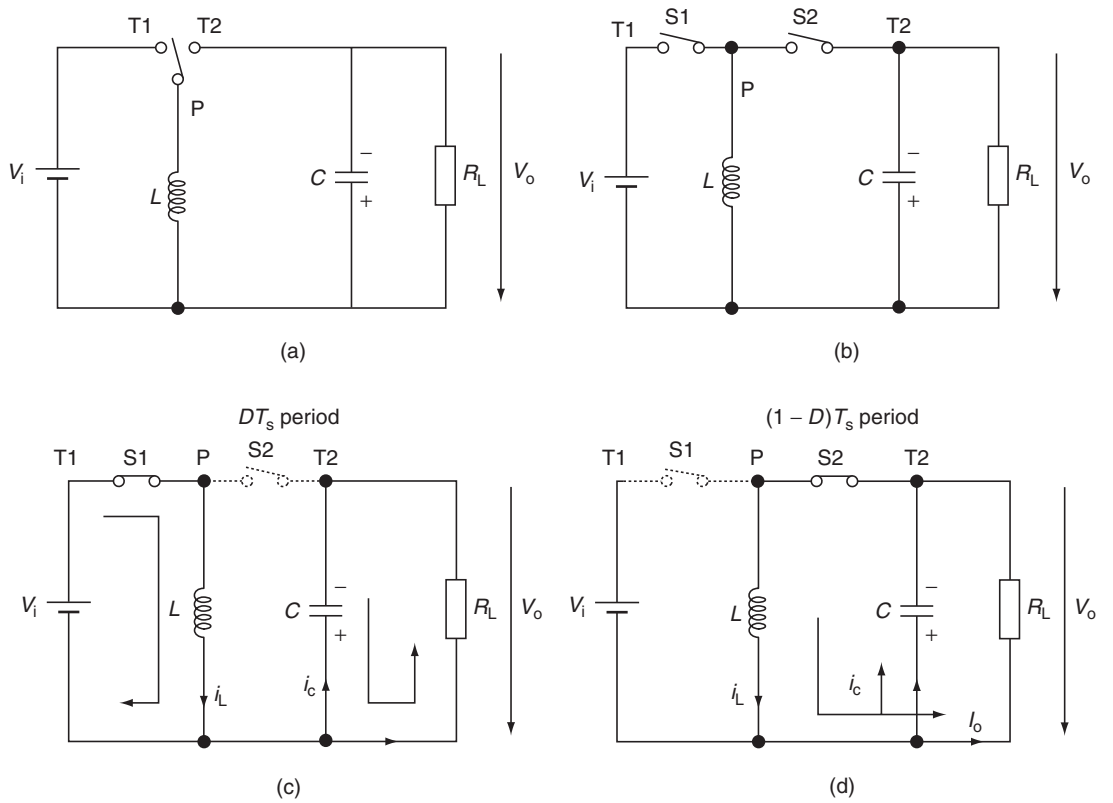


Figure 5.19 Step-up/down or buck–boost converter: (a) Buck–boost converter with a SPDT switch, an L and a C; (b) SPDT switch is replaced with two switches S1 and S2; (c) operative circuit during the period DT_s; (d) operative circuit during the period $(1 - D)T_s$.

Operation

The operation of the buck–boost converter is analyzed by visualizing the waveforms of the various signals of the circuit. The following signals are considered for analysis:

1. voltage across the inductor;
2. current through the inductor;
3. load current;
4. current through the capacitor;
5. input source current;
6. voltages across switches S1 and S2;
7. currents through switches S1 and S2.

Figure 5.20 shows the plot of the various signals of the buck-boost converter under steady-state operating conditions.

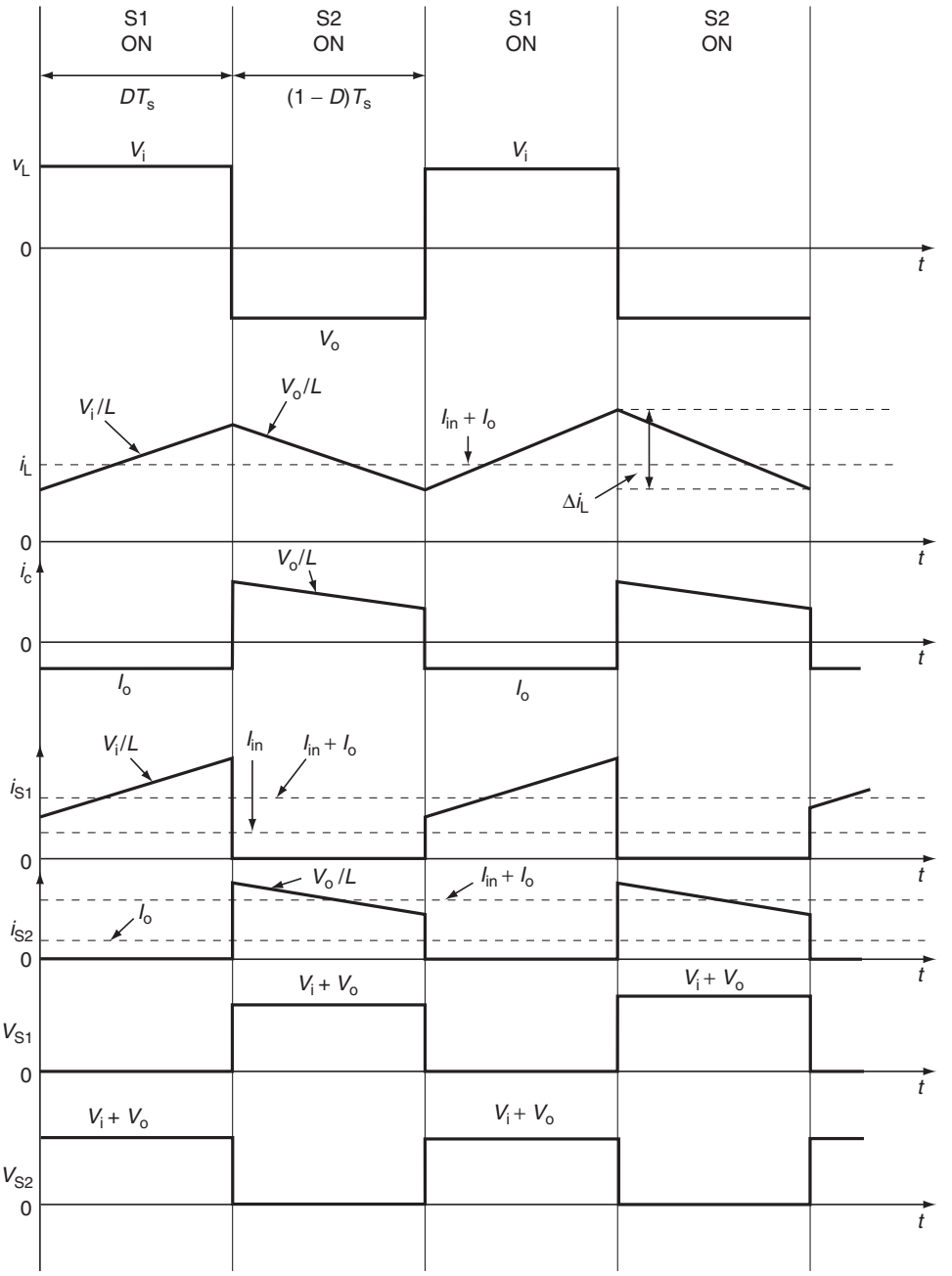


Figure 5.20 | Waveforms of various signals of a buck–boost converter.

DT$_s$ Duration The switch S1 is ON and S2 is OFF during the period DT_s. The voltage v_L across the inductor during this time is V_i which is a DC quantity. From the Faraday's law, the inductor voltage and current are related as $v_L = L(\mathrm{d}i_L/\mathrm{d}t)$. This implies that the current through the inductor is given as

$$i_L = \frac{1}{L}\int v_L \mathrm{d}t = \frac{1}{L}\int V_i \mathrm{d}t = \frac{V_i}{L}\int \mathrm{d}t$$

As V_i is a constant, the inductor current i_L rises linearly with a slope of V_i/L as indicated in Figure 5.20. During this time the current through S1 is the same as the inductor current. The voltage across S2, which is OFF, is $V_i + V_o$ by applying the Kirchhoff's voltage law.

(1 – D)T$_s$ Duration During the time when S1 is OFF and S2 is ON, the voltage v_L across the inductor during this time is V_o. However, V_o is a negative quantity with respect to the circuit ground. The inductor current i_L falls linearly with a slope of V_o/L as indicated in Figure 5.20. During this time, the current through S2 is the same as the current through the inductor. The voltage across S1, which is OFF, is $V_i + V_o$. The S1 switch current (i_{S1}) is shown in Figure 5.20. The average or DC component of the S1 switch current is I_{in}. The S2 switch current (i_{S2}) is also shown in Figure 5.20. The average or DC component of the S2 switch current is I_o. The i_{S2} has two distinct components as shown in Figure 5.20: (a) the average or DC component (I_o) and (b) the AC component (i_c). The DC component cannot flow through the capacitor; therefore it flows into the load R_L. The AC component alone will flow through the capacitor. Therefore, the capacitor current i_c is the S2 switch current minus the load DC component. The inductor current is composed of two parts, viz., (a) the average or the DC component that is the sum of the DC components of S1 and S2 switch currents (i.e., $I_{in} + I_o$) and (b) the AC component Δi_L. Here also the inductor current ripple is a design choice.

Semiconductor Switches

The boost converter schematic of Figure 5.19 shows the conceptual SPDT and SPST switches. The SPST switches may be replaced with power semiconductor switches as illustrated in Figure 5.21. Referring to the operating modes given in Figures 5.19 (c) and (d) and the OFF-state voltages across S1 and S2 given in Figure 5.20, S1 can be either a BJT or a MOSFET or an IGBT and S2 is a diode that turns ON when S1 is switched OFF by virtue of the inductor voltage ($L\mathrm{d}i/\mathrm{d}t$) build up and allows the inductor current to freewheel by charging the capacitor. This implementation is shown in Figure 5.21(a). However, the power semiconductor switches like MOSFETs and IGBTs are available along with their body diodes built-in; therefore, the switch configuration of Figure 5.21(b) is also common. Here the IGBT of S1 and the diode of S2 are the operative components of each switch in this topology.

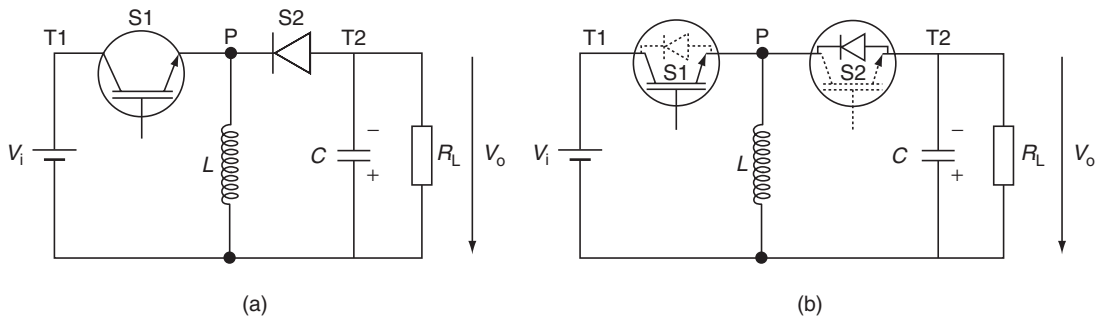


Figure 5.21 | Realization of buck–boost converter with power semiconductor switches.

Steady-State Analysis

Based on the application of the volt-second balance and charge balance rules, the input–output relationship under steady-state conditions can be obtained.

With Ideality Constraints The analysis is performed by assuming the following:

1. The switches S1 and S2 have no ON-state drops.
2. The winding resistance of the inductor is zero.
3. The ESR of the capacitor is zero.
4. There are no losses in the converter implying that the efficiency is 100%.

Input–output relationship To obtain the input–output voltage relationship, apply the volt-second balance rule to the inductor. This implies that the area under the inductor voltage curve in one period under steady-state conditions should be zero. Referring to Figure 5.20,

$$V_{\mathrm{i}} D T_{\mathrm{s}} + V_{\mathrm{o}} (1 - D) T_{\mathrm{s}} = 0$$

Simplifying the above equation, one obtains

$$V_{\mathrm{o}} = \frac{-V_{\mathrm{i}} D}{1 - D} \tag{5.22}$$

Note the negative sign in Eq. (5.22). This accounts for V_{o} being negative with respect to the circuit ground. The duty ratio D can take on values between the closed interval 0 and 1. If D is between 0 and 0.5, then the output V_{o} is less than V_{i} and results in buck operation. If it is between 0.5 and 1, then V_{o} is greater than V_{i} and results in boost operation. *Hence this topology is called a step-up/down converter or buck–boost converter*.

Note that V_{o} is infinite when D is unity. The same argument as discussed for the boost converter is valid here also. When $D = 0.5$, the converter acts as a unit pass circuit wherein $V_{\mathrm{o}} = V_{\mathrm{i}}$.

To obtain the input–output current relationship one can use the 100% efficiency constraint. This implies that

$$V_{\mathrm{i}} I_{\mathrm{in}} = \left| V_{\mathrm{o}} \right| I_{\mathrm{o}} \tag{5.23}$$

Using Eq. (5.22) in Eq. (5.23), one obtains

$$I_{\mathrm{in}} = I_{\mathrm{o}} \left(\frac{D}{1 - D} \right) \tag{5.24}$$

Inductor value The value of the inductor L is calculated based on the amount of current ripple Δi_{L} that the designer would like to allow for a given application. Referring to the waveforms given in Figure 5.20, the slope of the inductor current waveform during the period DT_{s} is given as

$$\frac{\mathrm{d}i_{\mathrm{L}}}{\mathrm{d}t} = \frac{\Delta i_{\mathrm{L}}}{\Delta t} = \frac{\Delta i_{\mathrm{L}}}{D T_{\mathrm{s}}} = \frac{V_{\mathrm{i}}}{L} \tag{5.25}$$

Likewise the slope of the inductor current waveform during the period $(1 - D) T_{\mathrm{s}}$ is given as

$$\frac{\mathrm{d}i_{\mathrm{L}}}{\mathrm{d}t} = \frac{\Delta i_{\mathrm{L}}}{\Delta t} = \frac{\Delta i_{\mathrm{L}}}{(1 - D) T_{\mathrm{s}}} = \frac{V_{\mathrm{o}}}{L} \tag{5.26}$$

V_i and V_o are known from the converter specifications. D can be obtained from the input–output relationship of Eq. (5.22). The switching frequency f_s is a design choice. From this $T_s = 1/f_s$ is obtained. Δi_L is also a design choice. Though this is chosen based on the application, the default choice is $\Delta i_L = 10\%$ of $I_{in} + I_o$ for the buck–boost converter. The output power P_o is known from the converter specifications. From the knowledge of P_o and V_o, I_o is calculated. Using Eq. (5.24), I_{in} can be estimated for a given load. The only unknown L can then be calculated from either Eq. (5.25) or Eq. (5.26). From Eq. (5.26), the value of inductor is given as

$$L = \frac{V_o(1-D)}{\Delta i_L f_s} \tag{5.27}$$

To regulate V_o with variations in V_i, the value of D will be varied. Thus if V_{imax} is the maximum input voltage swing, then D_{min} will be the corresponding minimum duty cycle to obtain a specified V_o. If V_{imin} is the minimum input voltage swing, then D_{max} will be the corresponding maximum duty cycle to obtain the specified V_o. Thus, for a regulated V_o,

$$V_o = \frac{V_{imax} D_{min}}{1-D_{min}} = \frac{V_{imin} D_{max}}{1-D_{max}} \tag{5.28}$$

V_o, V_{imax} and V_{imin} are known from specifications of the converter. D_{max} is a design choice that is dependent on the specific application. Then for the buck–boost converter,

$$D_{min} = \frac{V_{min} D_{max}}{V_{min} D_{max} + V_{max}(1-D_{max})}$$

The inductor value should be calculated for D_{min} to ensure that the inductor current ripple is within specified limits. Thus,

$$L = \frac{V_o(1-D_{min})}{\Delta i_L f_s} \tag{5.29}$$

Capacitor value The capacitor value is calculated by applying the amp-second rule. Referring to the capacitor current waveform given in Figure 5.20, if there is to be no charge build up or charge reduction in the capacitor, the area under the capacitor curve in one period should be zero. The area under the positive portion of the current curve implies charging of the capacitor and that under the negative portion implies discharging of the capacitor. For charge balance the area under both should be the same. The change is the capacitor charge ΔQ is given by the area under either the positive portion or negative portion of the capacitor current curve. For the buck–boost converter as in the case of the boost converter, it is easier to find the area under the negative portion of the capacitor current curve. Thus,

$$\Delta Q = C\Delta V_o = I_o D T_s = \frac{I_o D}{f_s} \tag{5.30}$$

$$C = \frac{I_o D}{\Delta V_o f_s} \tag{5.31}$$

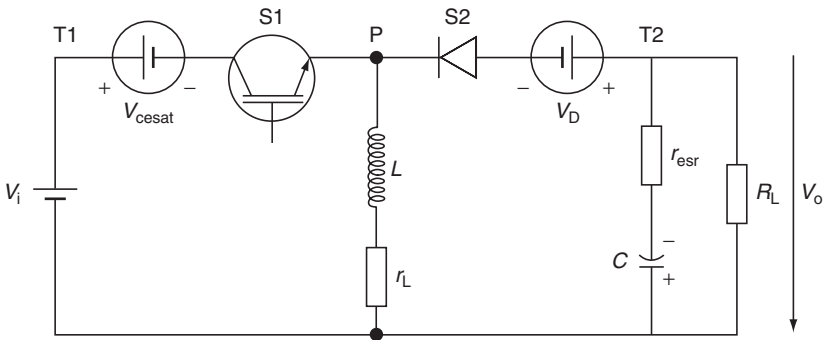


Figure 5.22 | Buck–boost converter with non-idealities included.

ΔV_o can be obtained from the output ripple specification. Δi_L and f_s are design choices as mentioned earlier. The value of C should be greater than the value calculated by Eq. (5.31).

Effect of Non-Idealities The analysis assumed that the switches, inductor and capacitor as ideal. However, practical components are far from ideal. As discussed for the buck and the boost converters, the switches have ON-state drops, the inductor has winding resistance, the capacitor has the ESR that contributes to output voltage ripple and reduction in efficiency. As a consequence, the output of the practical buck–boost converter will not exactly agree with the input–output relationship derived with the ideality constraints. Figure 5.22 shows a buck–boost converter schematic with few non-idealities included. The ON-state drops of the switches are included as voltage sources in series with the SPDT switches. The source V_{cesat} in series with the IGBT switch represents the ON-state drop of S1 when S1 is ON. The source V_D in series with the diode switch represents the ON-state drop of S2 when S2 is ON. The resistance r_L represents the inductor winding resistance and r_{esr} represents the capacitor ESR.

To derive a more exact input–output voltage relationship, one will have to apply the volt-second and amp-second balance to the circuit of Figure 5.22 wherein the non-idealities are included in the schematic. The design of the capacitor will also get affected due to the presence of the non-idealities especially the ESR non-ideality as discussed for the buck and boost converters.

5.5 Selection of Power Switches

Device Selection

The switch element in any converter is selected based on the ideal *i–v* < static characteristic requirement. The static *i–v* characteristic requirement is obtained from the operative converter circuits during the periods when the particular switch is ON and OFF. During the period when the switch is ON, the current flow requirement is obtained. During the period when the switch is OFF, the voltage withstanding capability requirement is obtained. The required idealized *i–v* characteristics is compared with the idealized *i–v* characteristics of the various power switching devices discussed in Chapter 1 and the most appropriate device type is chosen for the particular SPST switch position in the converter. The following example illustrates the selection of S1 and S2 switches.

EXAMPLE 5.1 **Buck Converter S1 and S2 Switch Selection**

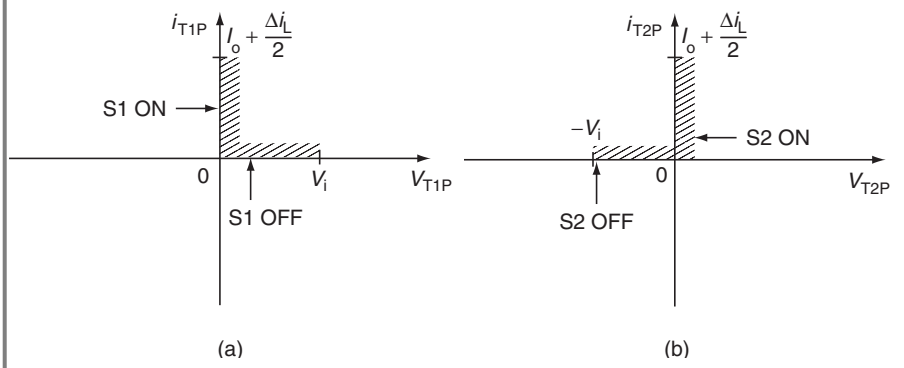


Figure 5.23 (a) S1 switch *i–v* requirement; (b) S2 switch *i–v* requirement for the buck converter.

Based on the operation of the buck converter and the waveforms shown in Figure 5.12, *i–v* requirements for S1 and S2 are drawn as shown in Figure 5.23. To be consistent for all switches and converters, *i–v* requirements should be plotted for switch currents flowing from the corresponding *throw* to the *pole* when the switch is ON. For voltage across the switch, the corresponding *throw* potential with respect to the *pole* should be considered. Figure 5.23(a) shows the *i–v* requirements for the S1 switch. When S1 is ON, a maximum current of $I_o + (\Delta i_L/2)$ will flow through it at zero drop across the device. When S1 is OFF, a maximum voltage drop of V_i occurs across it at zero current through it. If one compares these S1 *i–v* requirements with the idealized static *i–v* characteristics of devices discussed in Chapter 1, the BJT, MOSFET and IGBT with or without the internal body diode can handle the S1 requirement.

Likewise the S2 *i–v* requirement is shown in Figure 5.23(b). When S2 is ON, a maximum current of $I_o + (\Delta i_L/2)$ will flow through it at zero drop across the device. When S2 is OFF, a maximum voltage drop of $-V_i$ occurs at throw T2 with respect to the pole P. This *i–v* characteristic can be handled by a diode. Thus a diode is chosen for the S2 switch of the buck converter. However, in order to reduce the inventory of the components, the internal body diode of the MOSFET or IGBT itself may be used to perform the function of the S2 switch as shown in Figure 5.13(b). In such cases, both S1 and S2 will be the same type of power semiconductor devices; however, S1 will utilize the MOSFET or IGBT portion only and S2 will utilize the internal body diode portion only by giving zero gate drive to the MOSFET or IGBT portion. The selection of the power device for the boost, buck–boost or any converter can be done using the above approach.

Electrical Stress Rating

The electrical stresses that occur in the device are the voltage across the device and the current through the device. When ON, the current capability of the device should be higher than the maximum current that will pass through the device by a factor of 1.5 to 2. When OFF, the voltage withstanding capability of the device should be higher than the maximum voltage that occurs across the device by a factor of 1.5 to 2. The waveforms of voltage across the device and the current through the device will decide the voltage and current rating selection of the device.

Apart from the voltage and current ratings, the device will also encounter stressing during switching. The voltage across the device will swing from zero to maximum and vice-versa. Likewise the current through the device will also swing from maximum to zero and vice-versa. The rate of change of the voltage across switch and/or the current through the device has a bearing on the failure rate of the device. These rates should be less than the dv/dt and the di/dt ratings of the device, respectively. The switching frequency and the rise time and fall time required for the voltage across the device and the current through the device will decide the dv/dt and the di/dt ratings for the device.

Thermal Stress Rating

The operating junction temperature T_j of the power switching device is an important stress on the device that affects the failure rate of the device. It is directly dependent on the power dissipated within the device and the rate at which this dissipated power is conducted away to the ambient by means of heat sinks. The power dissipation in the device is due to (a) ON-state or conduction losses, (b) OFF-state losses and (c) switching losses. From the device voltage and current waveforms for the converter, the ON-state, the OFF-state and the switching losses are calculated as discussed for the various devices in Chapter 1. The sum of all these losses gives the total power dissipation in the device. This dissipated power is conducted away as heat to the ambient through the junction to case, case to heat sink and sink to ambient thermal resistances so as to maintain the required operating junction temperature below the maximum allowable junction temperature. The dissipated power, the ambient temperature and the required operating junction temperature will decide the thermal resistance of heat sink that needs to be selected. Heat sink sizing and selection are discussed in Chapter 13. The removal of heat from the junction and the thermal dynamics for the design of heat sinks is also discussed in detail later in Chapter 13.

One should note that all electronic components like the resistor, inductor and capacitor should be chosen such that they are within the electrical stress limits and also the thermal stress limits.

5.6 Continuous and Discontinuous Conduction Modes

Continuous conduction mode (CCM) and discontinuous conduction mode (DCM) are defined with respect to the inductor current. If the current through the inductor is continuous without becoming zero at any instant in a switching period, then the converter operation is referred to as the CCM. If the current through the inductor becomes zero for some portion of the switching period then the converter operation is referred to as the DCM. Referring to the inductor current waveforms of the three primary converters, it can be observed that the ripple portion of the current is a triangular waveform. As the switching period T_s is very small compared to the time constant associated with L and the load, the ripple slopes are considered constant. If the ripple slopes are constant, the Δi_L swing of the inductor current during either DT_s or $(1 - D)T_s$ periods is independent of the load I_o. However, the average value is dependent on the load I_o. Thus for a converter with input voltage, output voltage and frequency specified, the duty ratio D is determined. In such

a case Δi_L is of constant amplitude. This constant amplitude Δi_L will, however, swing up or down depending on the load I_o. However, it should be noted that, in a practical converter the output impedance R_o is not zero but finite. Therefore with change in I_o, there will be a change in V_o by an amount equivalent to the drop across the output impedance R_o. This will cause a change in the slope of the inductor current ripple as it is dependent on the converter output voltage.

Consider the case of the inductor current of a buck converter. Figure 5.24 shows the inductor currents for various load conditions. As the load current I_o is decreased, observe that the inductor current ripple amplitude Δi_L does not change; however, the average decreases. At a specific load, I_{omin}, the bottom of the inductor current ripple just touches the zero line. This is the boundary of CCM and DCM. If the load is decreased any further, there will be portions of the switching period when the inductor current will become zero and the converter is said to be operating in DCM.

Referring to Figure 5.24, at the boundary of CCM/DCM,

$$I_o = I_{omin} = \frac{\Delta i_L}{2} \tag{5.32}$$

The Eq. (5.32) implies that for the converter to operate in CCM, the condition $\Delta i_L / 2I_o < 1$ should be satisfied. Referring to Eq. (5.5), for a buck converter, the inductor current ripple is given as

$$\Delta i_L = \frac{V_o(1-D)T_s}{L} \tag{5.32a}$$

Dividing Eq. (5.32a) by $2I_o$, one obtains

$$\frac{\Delta i_L}{2I_o} = \left(\frac{V_o}{I_o}\right)\frac{(1-D)T_s}{2L} = \frac{1-D}{2L/R_L T_s} = \frac{1-D}{K}$$

Thus,

$$\frac{\Delta i_L}{2I_o} = \frac{1-D}{K} \tag{5.32b}$$

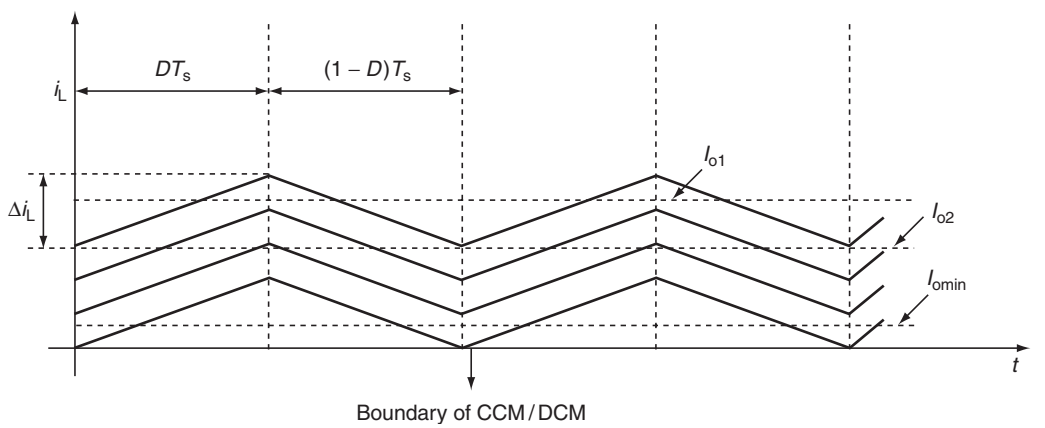


Figure 5.24 | Inductor current waveform for the buck converter with different loads.

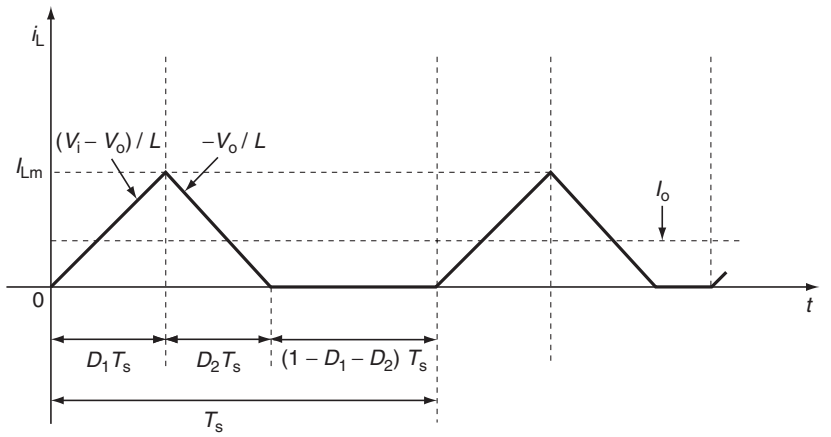


Figure 5.25 | Inductor current waveform in DCM operation for buck converter.

where $K \triangleq 2L / R_L T_s$ and is called the conduction parameter. From Eq. (5.32b), it can be seen that if $K > 1 - D$ then the buck converter is in CCM operation; if $K < 1 - D$ then the buck converter is in DCM operation and if $K = 1 - D$ then the buck converter is operating at the boundary of CCM and DCM. The value of K that results in the buck converter operating in the CCM/DCM boundary is called K_{bound}. Thus, $K_{bound} = 1 - D$ for the buck converter.

To maintain the converter operation in CCM, the load should not be decreased any lower than that given by Eq. (5.32). If the converter is designed for CCM operation, it should not be operated in DCM and vice-versa because the input–output relations and design equations will not be the same for both the modes of operations. Considering the buck converter, if the converter is operating in DCM, then in one switching period, there are three parts as indicated in Figure 5.25, viz., (a) duration $D_1 T_s$ when switch S1 is ON and the inductor current is rising, (b) duration $D_2 T_s$ when the switch S2 is ON and the inductor current is freewheeling and (c) duration $(1 - D_1 - D_2) T_s$ wherein the inductor current is zero. To obtain the input–output relationship, the volt-second balance rule should be applied to the voltage across the inductor. Thus,

$$(V_i - V_o) D_1 T_s + (-V_o)(D_2 T_s) + 0(1 - D_1 - D_2) T_s = 0$$

Solving the above equation, one obtains

$$V_o = V_i \frac{D_1}{D_1 + D_2} \tag{5.33}$$

The form of the Eq. (5.33) obtained above is not very useful for design purposes as there is one equation and two unknowns, that is, D_1 and D_2, V_o and V_i being known from the specifications of the converter. Therefore, the above equation should be further simplified such that there remains only one unknown parameter, either D_1 or D_2.

Referring to Figure 5.25, the peak inductor current I_{Lm} is used to relate D_1 and D_2. Thus using the rising and falling slope information of the inductor current waveform, one has

$$I_{Lm} = \left(\frac{V_i - V_o}{L} \right) D_1 T_s = \left(\frac{V_o}{L} \right) D_2 T_s \tag{5.34}$$

The average value of the inductor current I_o is given by

$$I_o = \frac{1}{T_s}\int_0^T i_L \, dt = \frac{\text{Area under triangle}}{T_s} = \frac{1}{2} I_{Lm}(D_1 + D_2) \tag{5.35}$$

From Eqs. (5.34) and (5.35), one obtains

$$I_o = \frac{V_o}{R_L} = \frac{1}{2} I_{Lm}(D_1 + D_2) = \left(\frac{V_i - V_o}{2L}\right) D_1 (D_1 + D_2) T_s \tag{5.36}$$

Equation (5.36) can be re-arranged to obtain

$$D_2 = \left(\frac{2L}{R_L T_s}\right)\left(\frac{V_o}{V_i - V_o}\right)\left(\frac{1}{D_1}\right) - D_1$$

From the definition of the conduction parameter K given in Eq. (5.32b), D_2 is now simplified as

$$D_2 = K\left(\frac{V_o}{V_i - V_o}\right)\left(\frac{1}{D_1}\right) - D_1 \tag{5.37}$$

In Eq. (5.37), V_i and V_o are known from the buck converter specification, the conduction parameter K is a design choice which gives a measure of the depth of DCM operation. Smaller the value of K, larger is the zero inductor current portion and vice-versa. Therefore for a given load, I_o, larger value of K would imply smaller peak inductor current I_{Lm}. However, one must note that compared to the CCM operation wherein only a part of the energy in the core of the inductor is transferred to the output, in the DCM operation, the complete inductor energy is transferred to the output. This will reduce the inductor core sizes for a specified amount of energy transfer every switching cycle. Herein lies the classic trade between inductor peak current and the inductor core size which dictates the choice of K for a particular application. Considering Eqs. (5.33) and (5.37), there are two equations and two unknowns. Thereby D_1 and D_2 can be determined.

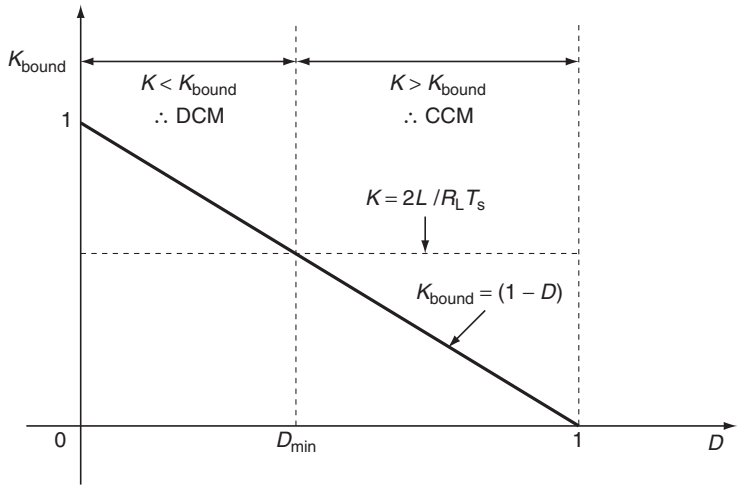


Figure 5.26 | Illustration of the choice of conduction parameter K.

As discussed earlier, K_{bound} is the value of the conduction parameter wherein the converter is at the boundary of CCM/DCM. The K_{bound} for the buck converter is shown to be $1 - D$. Figure 5.26 shows K_{bound} versus D. For a given load, the value of the conduction parameter K is also shown in Figure 5.26. For the range of D where $K > K_{bound}$, the converter operates in CCM and for the range of D where $K < K_{bound}$, the converter operates in DCM. The choice of K determines the range of duty ratio D for which the converter will operate in CCM and/or DCM. At the boundary of CCM/DCM, $K = K_{bound}$ and the corresponding duty cycle is the minimum duty ratio D_{min} for CCM operation of the converter.

Similar analysis can be performed for the other converters also. For the boost converter it can be shown that K_{bound} is $D(1 - D)^2$ and for the buck–boost converter the K_{bound} can be shown to be $(1 - D)^2$.

5.7 Isolated Converters

The three primary converters discussed till now do not provide isolation between the input source and the output load. However, in practical applications the power supplies and power converters are designed and built to incorporate isolation between the input source side circuit and the output load side circuit. Providing isolation has many benefits. Few of the important benefits are the following:

1. Galvanic isolation between source side and load side, that is, no physical connection between source side and load side. This suppresses all common mode electromagnetic interference (EMI).
2. The turns ratio can be used as a fixed gain for achieving output voltage amplification or attenuation. This will give the designer more flexibility to choose the range of duty ratio.
3. With isolation, the converter can be designed to have multiple outputs each with different voltages and grounds.

The isolated converters are topological extensions of the basic converters. Of the three primary converters, the buck and the buck–boost converters are extended to isolated topologies. The boost converter, however, does not lend itself well to isolated extensions. Thus the following two categories of isolated converters are popular:

1. buck-based isolated converters;
2. buck–boost-based isolated converter or flyback converter.

In the buck-based isolated converters there are many topological variants existing in the literature for different applications. However, the buck–boost-based isolated converter, or the flyback converter as it is known in the literature, is a direct extension of the buck–boost topology by including an isolating secondary winding on the same inductor core. The primary winding is connected to the source side circuit and the secondary winding is connected to the output side circuit. This will be discussed in detail later.

As there are many variants of the buck-based isolated converters, the schematic shown in Figure 5.27 gives a generic topology of the buck-based isolated converters. The buck-based isolated topologies can be apportioned into three parts as shown in Figure 5.27: (a) the output buck converter, (b) the input side chopper and (c) the isolation transformer.

The DC input source V_i cannot be directly fed to the isolation transformer as the transformer will function only if its primary is an AC voltage. Therefore, a chopper circuit is needed to convert the DC input source to an AC pulse voltage V_p before feeding to the transformer. Depending on the chopper topology, V_p may have a peak value of V_i or $V_i/2$. The secondary of the transformer will produce a voltage of nV_p as shown in the Figure 5.27. The value of the turns ratio n will determine whether nV_p will be amplified

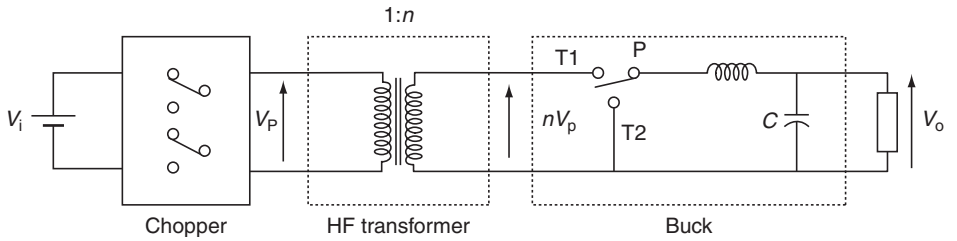


Figure 5.27 | Generic schematic of buck-based isolated converter.

or attenuated. The voltage nV_p is fed as input to the conventional buck converter. If the buck converter is switching with a duty ratio of D, then the output is given as

$$V_o = nV_pD \tag{5.38}$$

Based on the input chopper topology, the buck-based isolated converters are classified as follows:

1. **Forward Converter:** The chopper is a single unidirectional power flow SPDT switch.
2. **Push–Pull Converter:** The chopper consists of two back-to-back forward converters with the transformers wound on the same core.
3. **Half-Bridge Converter:** The chopper is a single bi-directional power flow SPDT switch.
4. **Full-Bridge Converter:** The chopper consists of dual bi-directional power flow SPDT switches.

Both the buck-based converter variants and the buck–boost-based flyback converter will be discussed in the following sections.

Forward Converter

Topology

The topology of the forward converter is shown in Figure 5.28. As the power flow from the input source to the output through the transformer is only in one direction (i.e., the forward direction from source to load), the converter is called the forward converter. It consists of a high-frequency transformer with primary-side chopper having an SPDT switch with pole P_p. The secondary side of the high-frequency transformer is the classical buck converter topology having the SPDT switch with pole P_s. The primary-side SPDT switch pole P_p can either connect to the negative terminal of the input source V_i or to the resistor R_f.

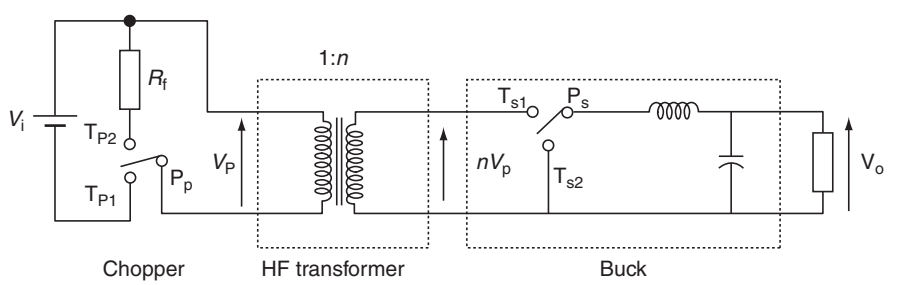


Figure 5.28 | Forward converter schematic.

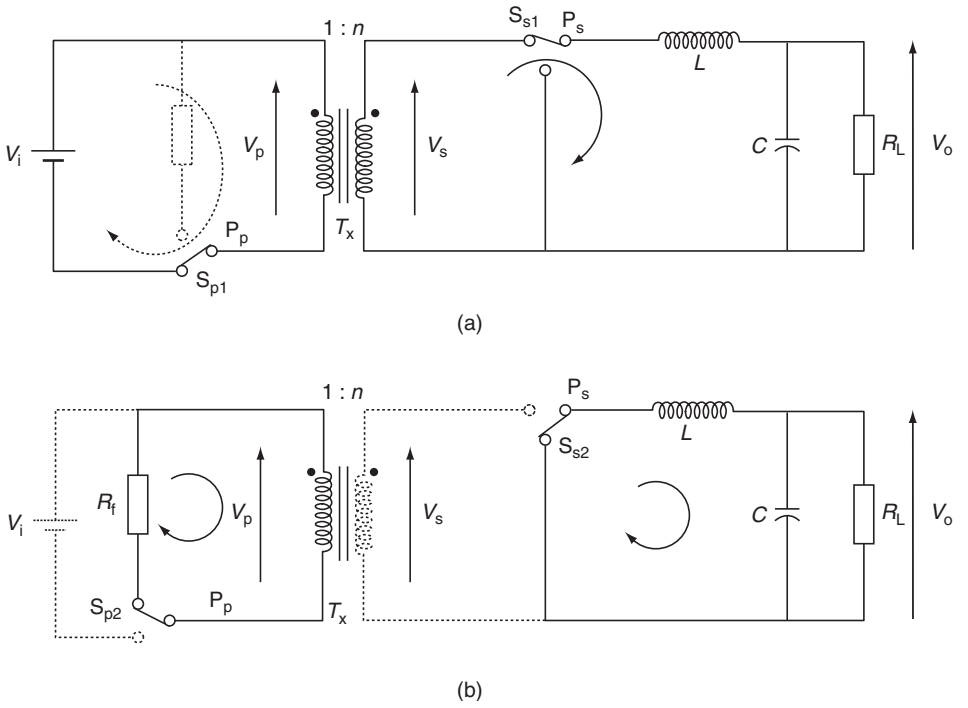


Figure 5.29 (a) Forward converter operative circuit during DT_s; (b) operative circuit during $(1 - D)T_s$.

The dot polarities of the transformer T_x are as shown in Figure 5.29. The operation of the forward converter is depicted in Figure 5.29. Figure 5.29(a) gives the operative circuit during the DT_s period and Figure 5.29(b) shows the operative circuit during the $(1 - D)T_s$ period.

DT_s Period

During this period the pole P_p of the primary-side SPDT switch is connected to the negative terminal of the input source V_i. The pole P_s of the secondary side SPDT switch is connected to the secondary winding through switch S_{s1}. The voltage across the primary is V_i during this time. The secondary voltage v_s is nV_i which is applied to the input of the secondary side buck converter. The power flows from the input source V_i through the transformer and into the buck converter. During this period, a constant voltage V_i is applied across the transformer primary. The flux in the core is governed by the equation,

$$v_p = L_m \frac{di_m}{dt} = N_p \frac{d\phi}{dt}$$

where L_m is the magnetizing inductance of the transformer, i_m the magnetizing current, N_p the primary number of turns and ϕ the flux in the core. This implies that the flux in the core is given by

$$\phi = \frac{V_i}{N_p} \int dt$$

As V_i and N_p are constants, ϕ increases linearly with time in this period.

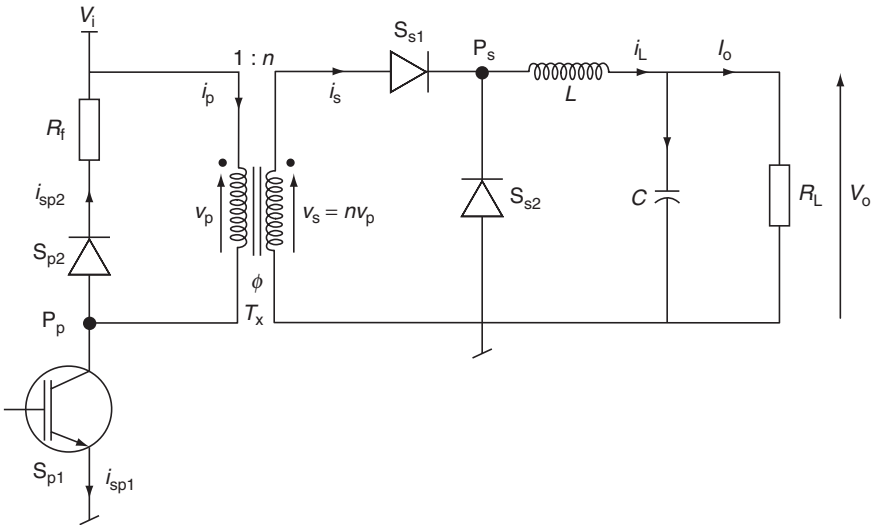


Figure 5.30 | Schematic of forward converter implementation.

(1 – D)T_s Period

During this period, the pole P_p is connected to the resistor R_f and the pole P_s is connected to the secondary side ground through switch S_{s2} allowing the inductor L to freewheel. As one of the secondary windings is open, there is no current flowing through the secondary of the transformer. When the primary side switch S_{p1} is switched OFF, the primary current is interrupted. The flux within the core is a continuous function of time and it will not become discontinuous. The sudden cut-off of the primary current implies a negative $\mathrm{d}i_m/\mathrm{d}t$ and therefore the voltage across the primary will reverse. This reverse voltage v_p that appears across the primary winding will tend towards infinity as i_m is cut-off instantaneously with the non-dot side of the transformer being more positive than the dot side. This high voltage will damage either the transformer insulation or the switch S_{p1}. Therefore, the pole P_p is connected to the resistor R_f through the switch S_{p2}. The negative voltage across the primary will cause a current to flow through R_f in the direction as shown in Figure 5.29(b). A negative voltage across the primary implies that $\mathrm{d}\phi/\mathrm{d}t$ is negative. This causes the flux in the core ϕ to decrease or decay with the time constant as decided by L_m/R_f.

Semiconductor Switches

The SPDT switches can be replaced by the power semiconductor switches by knowing the required *i–v* characteristics of each of the switches. This is shown in Figure 5.30. When switch S_{p1} is ON, the current through the switch flows from the pole to the throw. The voltage across the switch pole with respect to the throw is positive when S_{p1} is OFF. Therefore, a BJT, a MOSFET or an IGBT can be used for S_{p1}. When S_{p2} is ON, the current through the switch flows from the pole to the throw. The voltage across the switch pole with respect to the throw is negative when S_{p2} is OFF as T_{p2} is connected through R_f to V_i and T_{p1} is connected to ground. Thus a diode is used for S_{p2}. By similar arguments S_{s1} and S_{s2} are also replaced by diodes as shown based on their *i–v* requirements.

Waveforms

The waveforms of the various signals of the forward converter are shown in Figure 5.31. Each switching period is divided into two parts: (a) the DT_s period where the switches S_{p1} and S_{s1} are ON and the switches S_{p2} and S_{s2} are OFF and (b) the $(1 - D)T_s$ period where the switches S_{p2} and S_{s2} are ON and the switches S_{p1} and S_{s1} are OFF.

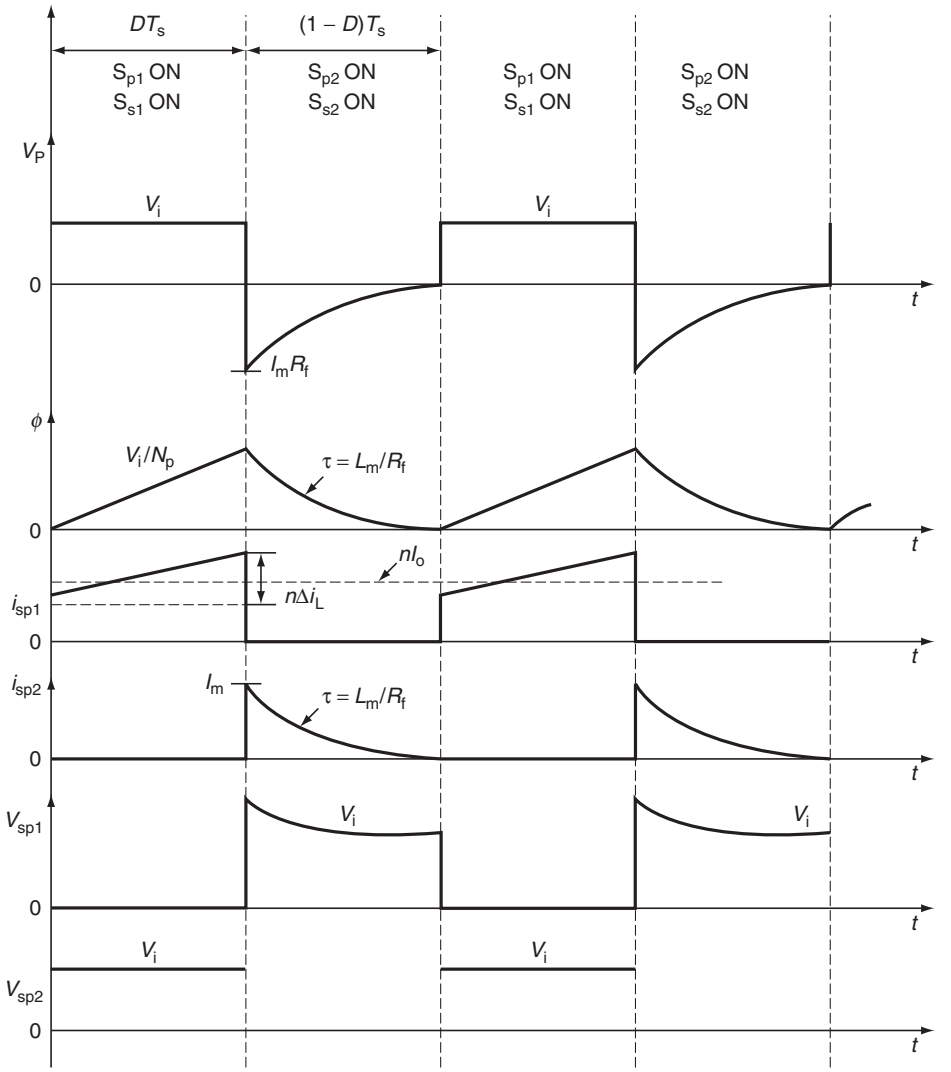


Figure 5.31 | Forward converter waveforms.

DT_s Duration

In this period the primary voltage v_p is equal to V_i the input source voltage. The secondary voltage v_s is equal to nV_i. As v_p is a constant, the flux in the transformer core ϕ is a linearly increasing function with slope V_i/N_p. As the pole P_p is connected to ground through switch S_{p1}, the voltage across S_{p2}, which is OFF, is V_i. The primary switch current i_{sp1} is the reflected secondary current. This is same as the reflected inductor current for this period. However, it should be noted that i_{sp1} also carries the magnetizing current i_m in addition to the load reflected component. The switch current i_{sp1} is shown enlarged in Figure 5.32. However, as I_m is generally very small compared to the load, it is usually neglected in design calculations.

$(1 - D)T_s$ Duration

During this period S_{s1} is OFF, therefore there is no secondary reflected current in the primary. The primary consists of only the decaying magnetizing current. The primary voltage v_p is decided by the voltage across R_f which is $i_m R_f$. Note that the dot of the T_x is negative with respect to the non-dot winding end as the magnetizing current freewheels through R_f. The magnetizing current i_m decays with the time constant of L_m/R_f. The flux waveform ϕ also decays with the same time constant. The primary switch S_{p1} is now OFF and will have to withstand a maximum voltage of $V_i + I_m R_f$ by applying Kirchhoff's voltage law to the input source side loop comprising V_i–R_f–S_{p1}.

Governing Equations

The operation of the output side buck converter has already been discussed earlier. The input–output voltage relationship is similar to the buck converter input–output relationship except for a factor of n which is the turns ratio of the transformer. The input–output relationship is given as

$$V_o = nV_i D$$

where $n = N_s/N_p$. The design of the transformer magnetics will be discussed in detail in Chapter 7. Here the design equations for the primary side chopper components will only be discussed. The primary side chopper consists of three components: (a) switch S_{p1}, (b) switch S_{p2} and (c) the freewheeling resistor R_f.

Switch S_{p1} Selection

Referring to Figure 5.32, the peak current through switch S_{p1}, I_{pm}, is given as

$$I_{pm} = nI_o + \frac{n\Delta i_L}{2} + I_m \tag{5.39a}$$

where I_{pm} is the peak primary current when switch S_{p1} is about to turn OFF. The peak current rating of the device, I_{cm}, should satisfy the following inequality:

$$I_{cm} > I_{pm} \tag{5.39b}$$

Referring to Figure 5.31, the device peak OFF-state voltage-handling capacity is given as

$$V_{ceo} > V_{imax} + I_m R_f + V_D \tag{5.40a}$$

where V_{imax} is the maximum input voltage, V_D is the diode drop of switch S_{p2} and I_m is the maximum magnetizing current.

However, it should be noted that Eq. (5.40a) is valid for an ideal transformer that has no leakage inductance reflected at the primary. In the case of a practical transformer, the coupling between the primary and

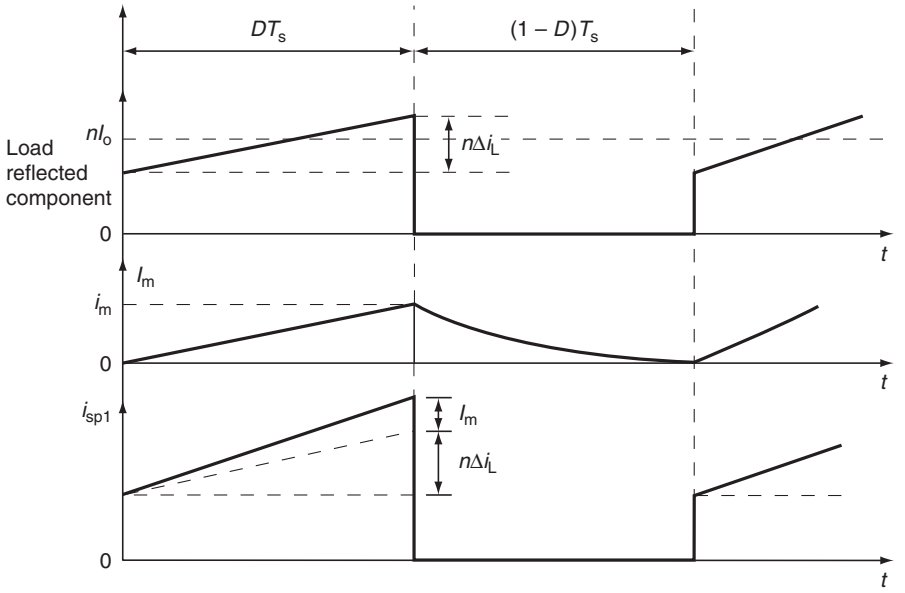


Figure 5.32 Primary switch current i_{sp1} showing the magnetizing and load reflected components.

the secondary windings is not 100%. There will be a finite leakage inductance as seen from the primary. As a consequence, when S_{p1} is ON, the leakage inductance L_σ stores an energy of $(1/2)L_\sigma I_{pm}^2$ when switch S_{p1} is about to turn OFF. This energy will also get dissipated in the resistor R_f. Thus, at turn-ON of switch S_{p1}, the voltage drop across R_f is $I_{pm}R_f$ rather than I_mR_f as in the ideal transformer case. Thus for a practical transformer with leakage inductance, the IGBT device peak OFF voltage-handling capacity should be

$$V_{ceo} > V_{imax} + I_{pm}R_f + V_D \tag{5.40b}$$

where I_{pm} is as given in Eq. (5.39a). The power dissipation of the switch S_{p1} should be calculated by evaluating the ON-state and switching losses as discussed in Chapter 1. This calculated power dissipation should be used to select the heat sink for the device to maintain the device junction temperature below the maximum specified value.

Switch S_{p2} Selection

The switch S_{p2} is a diode. When the diode is conducting, the peak current that it will encounter is I_m as can be observed from the waveforms of Figure 5.31. For the diode selection, it is important to estimate the peak, root mean square (rms) and the average currents flowing through it. The rms and the average currents are required to estimate the power dissipation in the diode which will facilitate the selection of the heat sinks if required. Thus,

$$I_{dm} > I_{pm} \tag{5.41}$$

where I_{dm} is the peak diode current rating. The load reflected portion of the primary current that is stored in the leakage inductance will decay quickly as the energy stored in the leakage inductance is small. Therefore,

for calculating the diode average and rms currents only I_m can be used. The diode conducts during the period $(1 - D)T_s$. The current waveform is an exponential decaying waveform. The average current rating of the diode should be greater than the average value of the exponentially decaying waveform as shown in Figure 5.31. However, it is difficult to find the average of the exponentially decaying waveform. Without loss of generality, the exponentially decaying waveform can be approximated to a triangular waveform that would enclose the exponentially decaying waveform entirely. For a triangular waveform approximation,

$$I_{\text{d-avg}} > \frac{1}{2} I_m (1 - D) \tag{5.42}$$

where $I_{\text{d-avg}}$ is the average current rating of the diode and

$$I_{\text{d-rms}} > I_m \sqrt{\frac{1 - D}{3}} \tag{5.43}$$

where $I_{\text{d-rms}}$ is the rms current rating of the diode. When the diode is in the OFF condition, the peak inverse voltage (PIV) across the diode is given as

$$PIV > V_i \tag{5.44}$$

Selection of Resistor R_f

The function of the resistor R_f is to dissipate the magnetic energy so that the flux in the core gets reset to zero. The energy stored in the inductor at the time of turn-OFF of switch S_{p1} is given as

$$\varepsilon_m = \frac{1}{2} L_m I_m^2 \tag{5.45}$$

This energy as well as the energy stored in the leakage inductance has to get dissipated in the resistor R_f before the next switching cycle begins. Thus, in every switching period T_s, an amount of energy ε_m as given by Eq. (5.45) will be dissipated in resistor R_f. Thus, the power dissipated in the resistor is given as

$$P_{Rf} = \frac{(1/2) L_m I_m^2 + (1/2) L_\sigma I_{pm}^2}{T_s} = \frac{1}{2} (L_m I_m^2 + L_\sigma I_{pm}^2) f_s \tag{5.46}$$

It may be noted from Eq. (5.46) that the power dissipated in the resistor R_f is independent of the value of the resistor. The resistor value will therefore only affect the time constant of the magnetizing current decay. The flux and the magnetizing current should decay to zero in the $(1 - D)T_s$ period before S_{p1} switches ON again. As the decay is exponential in nature, one can expect the flux to have decayed to nearly zero in about five time constants. Therefore, to ensure that the flux decays to zero during the time S_{p2} is ON, one should have

$$5 \frac{L_m}{R_f} < (1 - D_{max}) T_s \text{ as } L_m \gg L_\sigma$$

or

$$R_f > \frac{5 L_m f_s}{1 - D_{max}} \tag{5.47}$$

Another constraint for R_f comes from the voltage withstanding capability of the IGBT switch (i.e., switch S_{p1}). From Eq. (5.40b), one obtains

$$R_f < \frac{V_{ceo} - V_{imax} - V_D}{I_{pm}} \tag{5.48}$$

From Eqs. (5.47) and (5.48) the bounding values for the selection of the resistor value are obtained as

$$\frac{5L_m f_s}{1 - D_{max}} < R_f < \frac{V_{ceo} - V_{imax} - V_D}{I_{pm}} \tag{5.49}$$

In the case of transformer with negligible leakage inductance the above inequality reduces to

$$\frac{5L_m f_s}{1 - D_{max}} < R_f < \frac{V_{ceo} - V_{imax} - V_D}{I_m}$$

If the inequality of Eq. (5.49) is not satisfied, then the maximum duty cycle for the converter should be reduced and/or the V_{ceo} rating of the primary side switch should be increased till the inequality Eq. (5.49) is satisfied.

Forward Converter with Demagnetizing Winding

The forward converter topology of Figure 5.30 has two main drawbacks:

1. The flux resetting is lossy as the magnetizing energy is lost in the resistor R_f.
2. As the flux decays asymptotically to zero and does not become truly zero at the start of the next switching cycle, there exists a finite average flux which steadily builds up and saturates the core. This has an upper limit on the operational duty ratio of the converter to avoid flux build up.

To address the drawbacks mentioned above, the lossy flux resetting part of the forward converter of Figure 5.30 is modified by introducing the demagnetizing or the flux resetting winding as shown in Figure 5.33(a). The primary side chopper that includes the SPDT switch is re-positioned with the pole P_p at the primary side ground. One of the throws (switch S_{p1}) is connected to the primary winding and the other throw (switch S_{p2}) is connected to the demagnetizing winding as shown in Figure 5.33(a). Observe the dot polarity of the demagnetizing winding with respect to the primary winding as shown. It is an important part of the operation of demagnetizing circuit.

Circuit Operation

The operation of the forward converter is shown in Figures 5.33(b) and (c). During the time DT_s, the operation of the forward converter is exactly same as that of the forward converter of Figure 5.30. During this time when the switch S_{p1} is ON, the voltage across the primary is V_i with the dot end positive. The demagnetizing winding will have a voltage of $V_i(N_d/N_p)$ with the dot end positive. As S_{p2} is OFF, there is no current flowing in the demagnetizing winding and it is not operative during this period. The difference is in its operation during the time $(1 - D)T_s$. During this time, there is no change in the operation of output buck portion of the forward converter. The switch S_{p1} is OFF and the switch S_{p2} is ON as shown in Figure 5.33(c). Also, the primary current is cut-off suddenly when S_{p1} is switched OFF. This causes a negative di/dt that will reverse the primary voltage polarity making the non-dot end positive. Likewise, by induction, the non-dot end of the demagnetizing winding will be positive. This will cause a current to flow through the demagnetizing winding against the source voltage V_i as shown in the Figure 5.33(c). The voltage across the demagnetizing winding will be V_i which will cause the flux in the core to decay linearly with a slope of V_i/N_d.

To reduce the flux leakage and keep the leakage inductance seen from the primary to a minimum, N_d is made equal to N_p so that both the primary and the demagnetizing windings can be wound together in a bifilar fashion. However, if $N_d = N_p$, then for the flux that is built up to a maximum of ϕ_m during DT_s will decay to zero in exactly the same time when S_{p1} is switched OFF and S_{p2} is ON. Thus if $N_d = N_p$, then $(1 - D)T_s > DT_s$ to ensure that flux will decay to zero. From this inequality, it can be seen that $D < 0.5$, that is, the maximum duty cycle is limited to 50% if $N_d = N_p$.

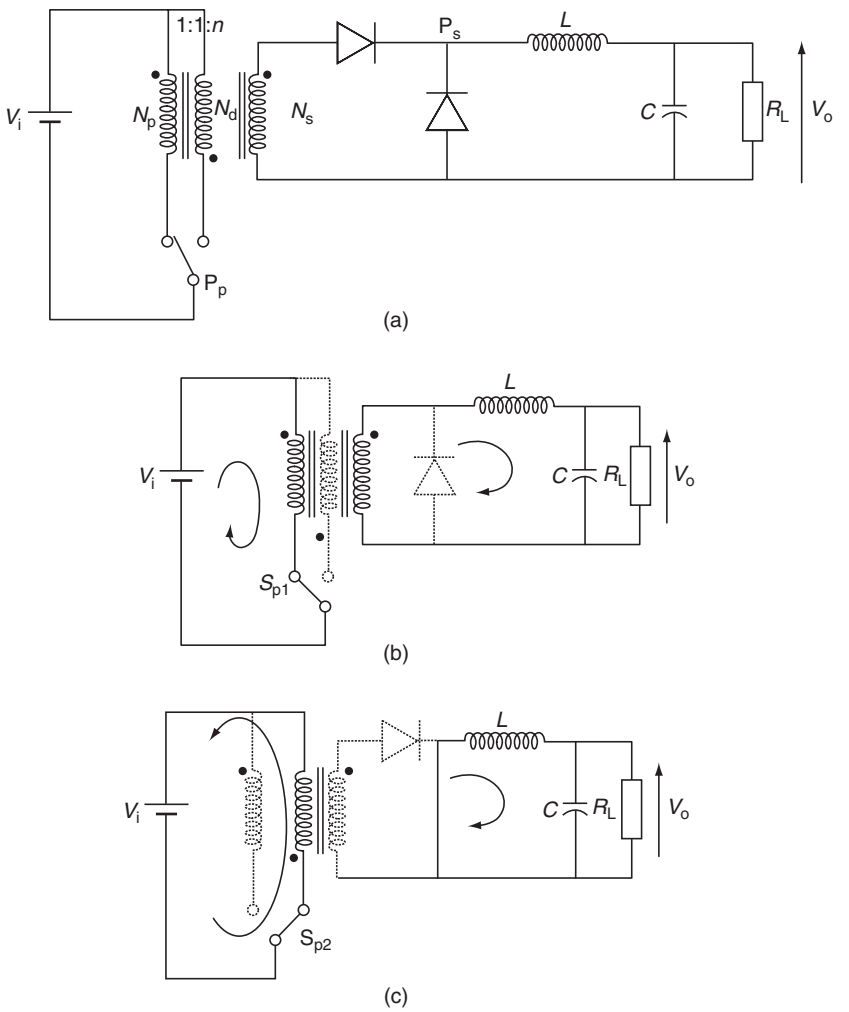


Figure 5.33 (a) Forward converter schematic with flux reset winding or demagnetizing winding; (b) operative circuit during DT_s; (c) operative circuit during $(1 - D)T_s$.

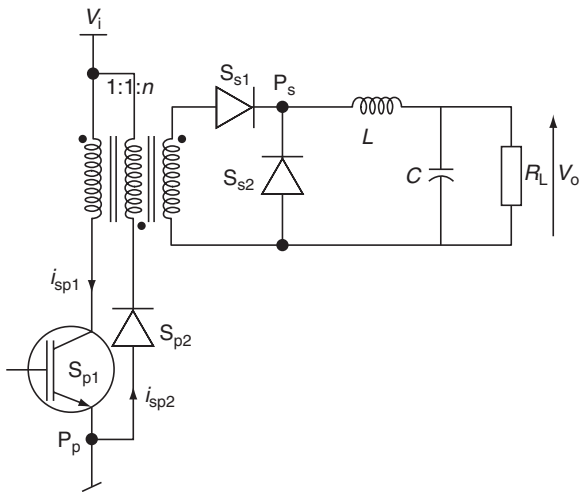


Figure 5.34 | Forward converter implementation circuit with demagnetizing winding.

Figure 5.34 shows the circuit schematic of the forward converter with demagnetizing winding. The SPDT switch is replaced with power semiconductor switches. S_{p1} is implemented with IGBT and S_{p2} is implemented with a diode. The governing equations for selecting the two switches are as follows.

Switch S_{p1} Selection There is no change in the device peak current rating I_{cm}. It is same as Eq. (5.40a) and is repeated here for easy reference:

$$I_{cm} > nI_o + \frac{n\Delta i_L}{2} + I_m$$

The device peak OFF-state voltage-handling capacity is given as

$$V_{ceo} > V_{imax} + (V_{imax} + V_D)\frac{N_p}{N_d} \tag{5.50}$$

where V_{imax} is the maximum input voltage, V_D the diode drop of switch S_{p2} and I_m the maximum magnetizing current. Observe that if V_D is considered negligible and $N_p = N_d$, then V_{ceo} should be greater than $2V_{imax}$. The power dissipation of the switch S_{p1} should be calculated by evaluating the ON-state and switching losses as discussed in Chapter 1. This calculated power dissipation should be used to select the heat sink for the device to maintain the device junction temperature below the maximum specified value.

Switch S_{p2} Selection The switch S_{p2} is a diode. When the diode is conducting, the peak magnetizing current that is reflected from the primary winding is $I_m(N_p / N_d)$. The peak, rms and average currents flowing through the diode are

$$I_{dm} > I_m \frac{N_p}{N_d} \tag{5.51}$$

where I_{dm} is the peak diode current rating. The diode conducts during the period $(1 - D)T_s$. The current waveform is a triangular waveform. Thus,

$$I_{d\text{-}avg} > \frac{1}{2}\frac{N_p}{N_d} I_m (1-D) \tag{5.52}$$

where $I_{d\text{-}avg}$ is the average current rating of the diode and

$$I_{d\text{-}rms} > \frac{N_p}{N_d} I_m \sqrt{\frac{1-D}{3}} \tag{5.53}$$

where $I_{d\text{-}rms}$ is the rms current rating of the diode.

When the diode is in the OFF condition, the *PIV* across the diode is given as

$$PIV > V_{imax} + V_{imax}\frac{N_d}{N_p} \tag{5.54}$$

Figure 5.35 shows the typical waveforms for the forward converter with demagnetizing winding. The waveforms are drawn for $N_p = N_d$. Observe that the flux decays at the rate of V_i/N_d. If the duty ratio D is less than 0.5, then the flux decays to zero much before $(1 - D)T_s$ time is completed. The voltage across the S_{p1} switch is $2V_i$ till the demagnetizing winding conducts. Once the flux has completely decayed to zero, the voltage across the primary winding is zero and only V_i appears across S_{p1} which is in OFF-state. The diode current (switch S_{p2}) is a triangular waveform wherein the magnetizing current decreases at the rate of V_i/L_m. When the switch S_{p2} is OFF during the period DT_s, it has to withstand a PIV of $2V_i$ as indicated.

Dual Switch Forward Converter

The forward converter with demagnetizing winding addressed both the issues of lossy flux reset and true zero flux reset value. However, there is one problem in this topology that can unduly stress the primary side switch S_{p1} which is the finite leakage inductance as seen from the primary. This issue is illustrated in Figure 5.36(a) where the leakage inductance between the primary and the demagnetizing winding as seen from the primary is shown as L_σ. The energy stored in the L_σ is not coupled to the demagnetizing side. Therefore, when S_{p1} is switched OFF, the energy in L_σ will cause a large reverse $L_\sigma(di_m/dt)$ across the leakage inductor. As i_m is cut-off instantaneously, the di_m/dt and hence the reverse voltage drop across the leakage inductor will be very large. This drop will appear across S_{p1} when it is being turned OFF and appear as the electrical stress on the device. This stress may damage the IGBT (switch S_{p1}). Even if N_p is made equal to N_d and both the windings are wound in bifilar fashion, there will be a finite leakage inductance as seen from the primary.

Figure 5.36(b) provides a solution to this problem. Instead of using a separate demagnetizing winding to do the function of flux resetting, the same primary winding is used to do the function of flux resetting. In this way there is no issue of magnetic coupling. The primary winding consists of two SPDT switches, one at each end of the winding. The SPDT switches comprise IGBTs and diodes as shown in Figure 5.36(b).

During the time DT_s, switches S1 and S2 are ON allowing power flow to the output buck converter module. At this time, the voltage across the primary is V_i with the dot end positive. During this time, the primary current flows from the input source through the path V_i–S1–A–B–S2–ground. During the time $(1 - D)T_s$, the switches S1 and S2 are OFF and the diodes D_1 and D_2 turn ON by virtue of the voltage reversal across the primary winding and leakage inductance L_σ in order to drive the magnetizing current through the diodes against the input voltage source V_i. At this time, the magnetizing current freewheels

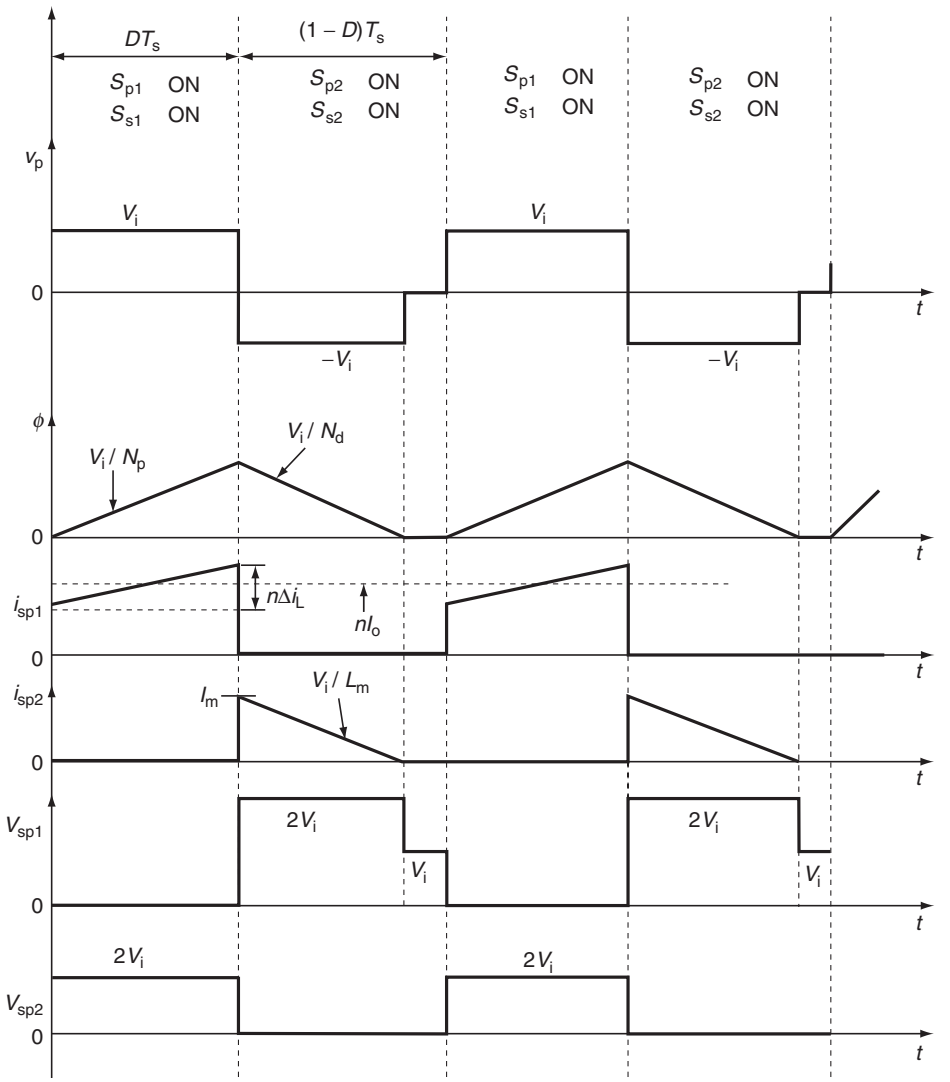


Figure 5.35 | Forward converter waveforms with demagnetizing winding.

through the path ground–D_1–A–B–D_2–V_i. Thus, during both the periods the current direction in the primary winding is maintained in the direction A to B. Observe that even the energy stored in the leakage inductance, as seen from the primary, freewheels through the diodes D_1 and D_2. As a consequence, the voltage spikes stress on the primary switches S1 and S2 is absent. During the time when S1 and S2 are OFF, D_1 and D_2 are conducting resulting in voltage across the OFF-state switches to be V_i. Likewise when S1 and S2 are ON, D_1 and D_2 will have to withstand voltage of V_i. Thus the voltage rating of the switches can be lower than that for the single switch forward converter topologies. All other governing equations for the switches are the same as discussed for the forward converter with demagnetizing winding.

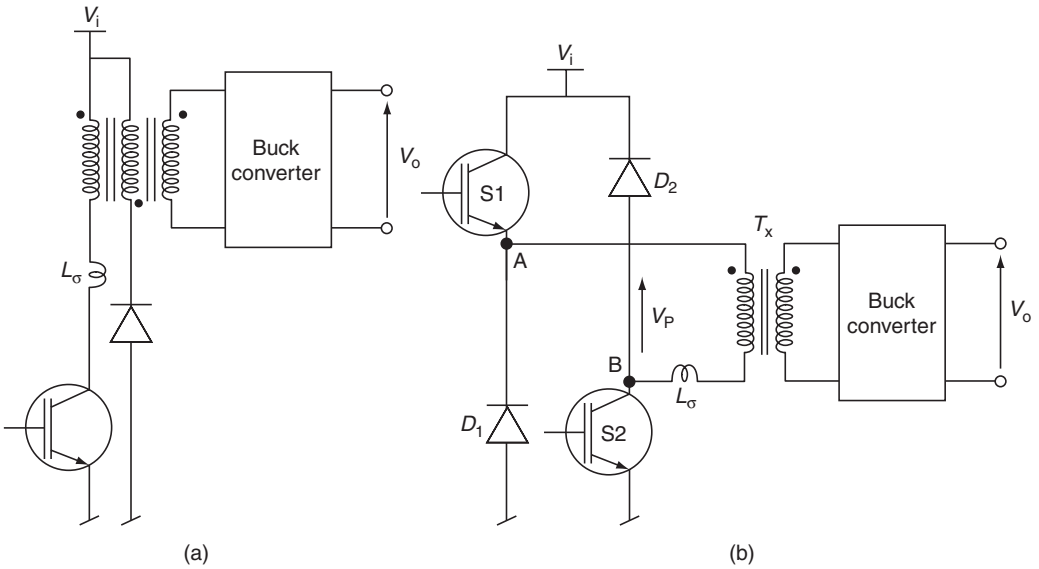


Figure 5.36 | (a) Leakage inductance problem; (b) dual switch forward converter.

Push–Pull Converter

The forward converter utilizes only the positive half of the core magnetization as the magnetizing current and the core flux are unidirectional. As a consequence the core is under-utilized and the size of the core for a given output power is larger than if core magnetization had been done in both directions. The push–pull converter magnetizes the core in both directions to better utilize the core. It is so named because it consists of two forward converters operating back to back. During the period DT_s, one forward converter is operative in transferring power to the load and during the period $(1 - D)T_s$ the other forward converter is operative in transferring power to the load. The schematic of the push–pull converter is shown in Figure 5.37.

It consists of a primary side single pole triple pole switch with pole P_p as shown. The transformer is center-tapped both in the primary and the secondary. The secondary side consists of the buck converter. The operating modes of the circuit are shown in Figure 5.38. Each part of the center-tapped winding behaves as a simple forward converter that comes into operation on alternate cycles. This is illustrated in Figure 5.38.

Figure 5.38(a) is the operative circuit when the pole P_p is connected to T_{p1}, that is, switch S_{p1} is ON as shown. The section V_i–C_1–B_1–S_{p1} acts as the primary portion of the forward converter. During this time the dot end is positive. Due to magnetic coupling, all the dot ends are positive with respect to their non-dot ends. As a consequence diode D_1 is forward-biased and diode D_2 is reverse-biased. The A_2–C_2 winding is the active winding during this period. It supplies the power to the buck converter module.

Figure 5.38(b) is the operative circuit when the pole P_p is connected to T_{p2}, that is, switch S_{p2} is ON as shown. The section V_i–C_1–A_1–S_{p2} acts as the primary portion of the forward converter now. During this time the non-dot ends are positive with respect to their dot ends. As a consequence, D_2 is forward-biased and D_1 is reverse-biased. The B_2–C_2 winding is the active winding which delivers power to the output during this time.

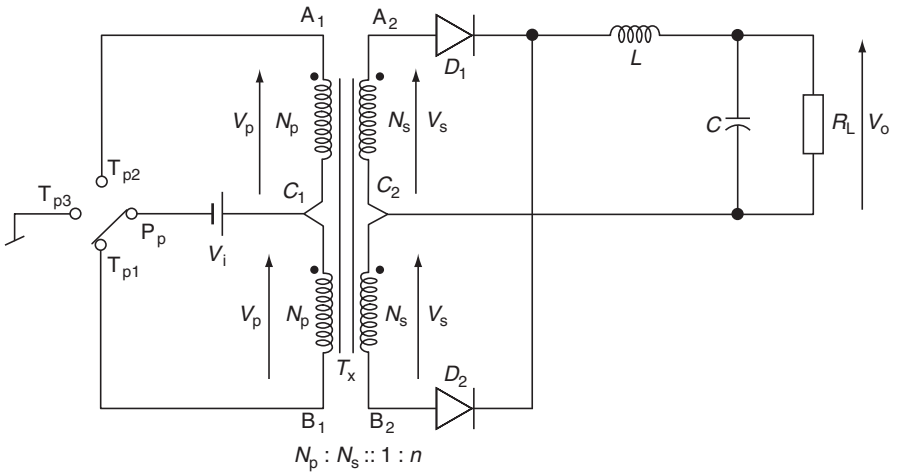


Figure 5.37 | The push–pull converter schematic.

Figure 5.38 shows the third operative mode of the push–pull circuit. Here the pole P_p is connected to the primary circuit ground as shown. Both switches S_{p1} and S_{p2} are OFF. During this time the inductor current freewheels by sharing the current between D_1 and D_2. Thus in the secondary, there is a current leaving the dot at A_2 and an equal current leaving the non-dot end at B_2. These two equal and opposite flux producing components will make $d\phi/dt$ in the core zero and thus the voltages across all windings are zero.

The waveforms for the various parameters of the push–pull circuit are shown in Figure 5.39. During the first DT_s period, S_{p1} is ON and S_{p2} is OFF. The voltage across the primary winding A_1 to B_1 is $2V_i$ and that across the secondary winding between A_2 and B_2 is $2nV_i$ with the dot ends being positive in all windings. The flux in the core increases linearly with a rate of V_i/N_p from $-\phi_m$ to $+\phi_m$. The diode D_1 carries the inductor current during this time. During this time the voltage across the OFF switch S_{p2} is $2V_i$.

During the $(1 - D)T_s$ period, both the primary switches S_{p1} and S_{p2} are OFF. As explained before, the $d\phi/dt$ is zero during this time. Thus ϕ is a constant in this time period. The voltage across the primary side OFF switches S_{p1} and S_{p2} is V_i as there are zero voltages across the windings. Observe also that the inductor current freewheels by equally sharing its current between the secondary diodes D_1 and D_2.

During the next DT_s period, S_{p2} is switched ON and S_{p1} is OFF. This makes the dot ends of the windings negative with respect to their non-dot ends, thereby reversing the flux to swing from $+\phi_m$ to $-\phi_m$ with a rate of $-V_i/N_p$. Every alternate cycle S_{p2} is switched ON to reverse the flux. It can be observed that the duty ratio D is defined with respect to the waveforms of the inductor current ripple. The charging up of the inductor energy is defined as the DT_s period and the freewheeling portion of the inductor current is defined as the $(1 - D)T_s$ period. However, the switches S_{p1} and S_{p2} are switched every alternate cycles and therefore, the switching frequency of the primary side switches is half the frequency of the output buck converter as can be seen from the waveforms of Figure 5.39.

The circuit schematic of the push–pull converter is shown in Figure 5.40 wherein the SPTT switch is replaced by two IGBTs with internal body diodes. The governing equations are:

$$V_o = nV_iD \tag{5.55}$$

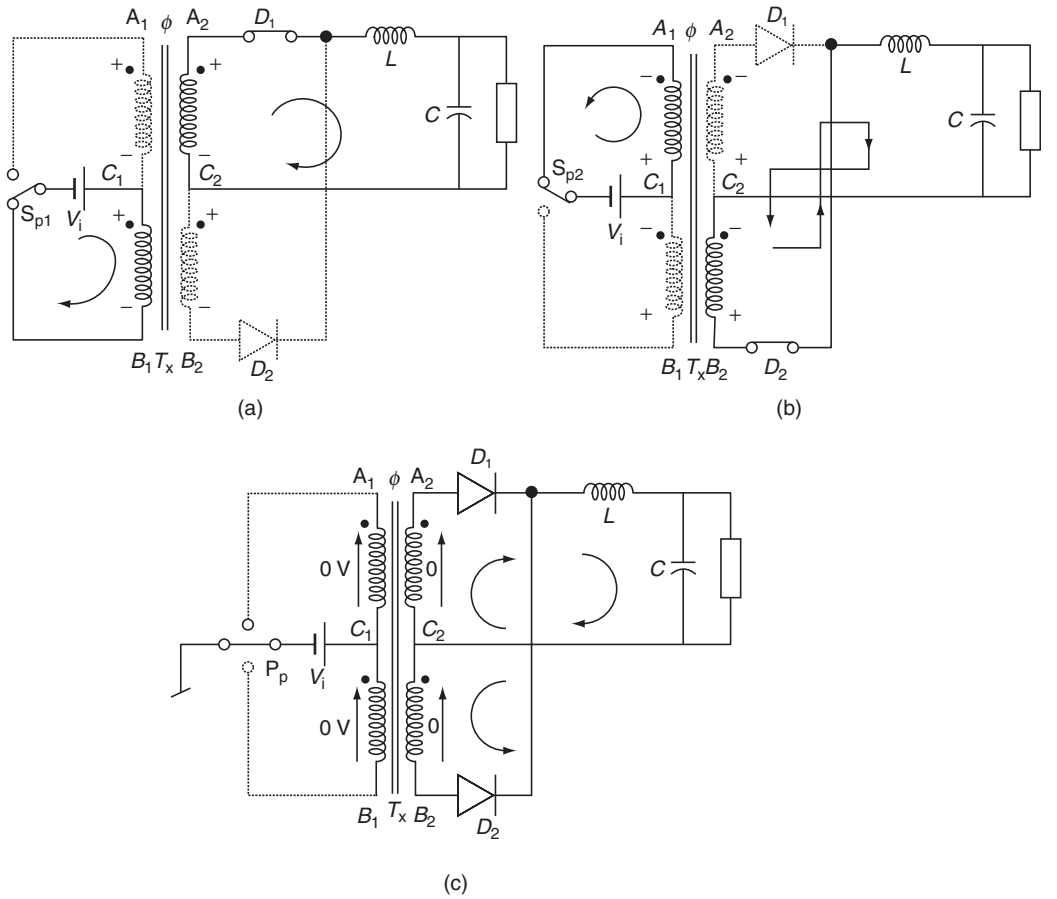


Figure 5.38 (a) Operative circuit during DT_s (S_{p1} ON); (b) operative circuit during DT_s (S_{p2} ON) of alternate cycle; (c) operative circuit during $(1 - D)T_s$ when both primary switches are OFF (i.e., S_{p1} and S_{p2} are OFF).

where n is the secondary to primary turns ratio and D is the duty ratio as defined for the buck converter inductor current waveform.

Primary Side Switch Selection

Referring to the waveforms of Figure 5.39, the current rating for the primary side switches S_{p1} and S_{p2} is

$$I_{cm} > nI_o + \frac{n\Delta i_L}{2} \tag{5.56}$$

where I_{cm} is the peak current requirement of the IGBT. Referring to Figure 5.39 again, the peak OFF-state voltage that the primary side switches should withstand is

$$V_{ceo} > 2V_{imax} \tag{5.57}$$

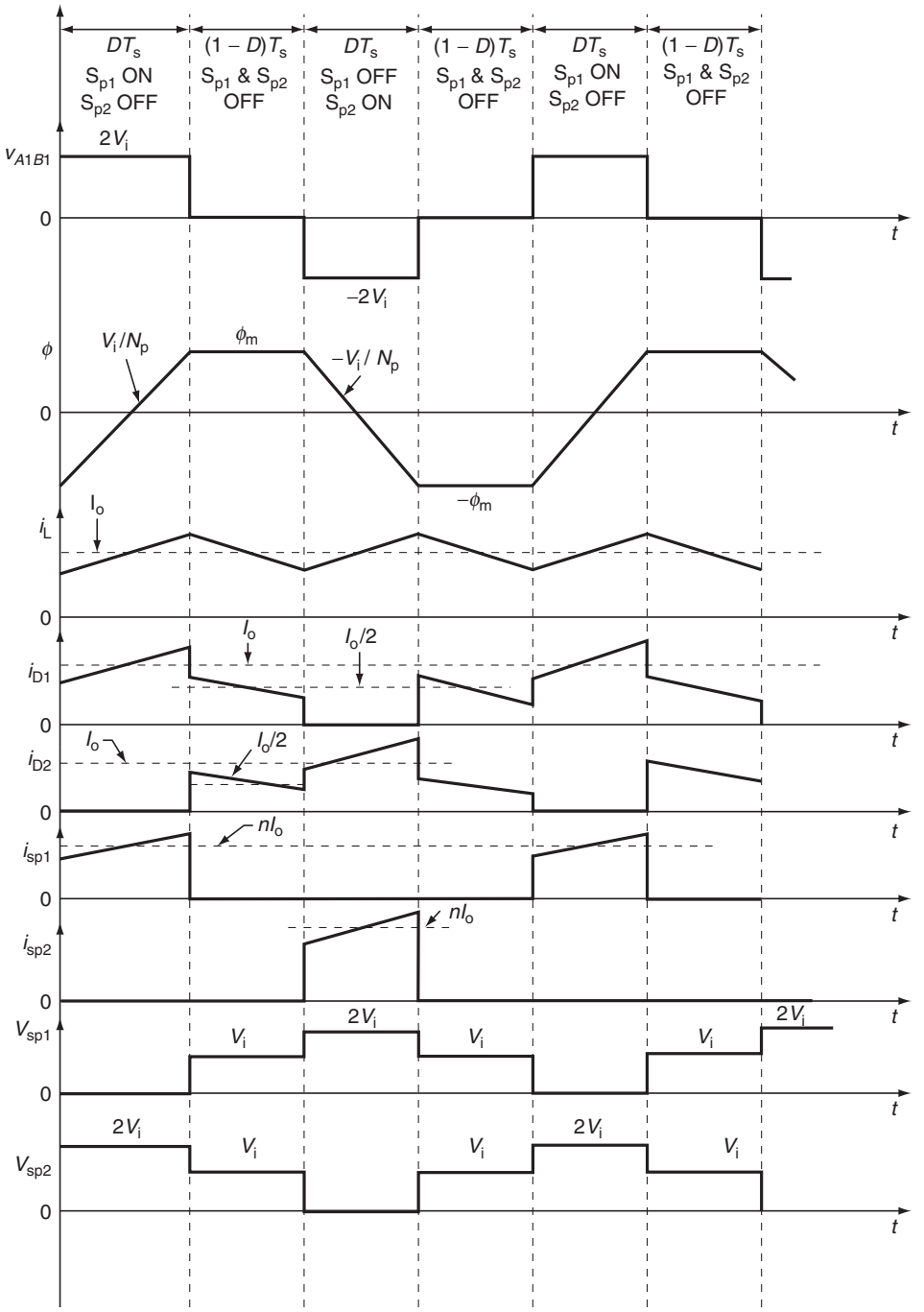


Figure 5.39 | Waveforms for the push–pull circuit.

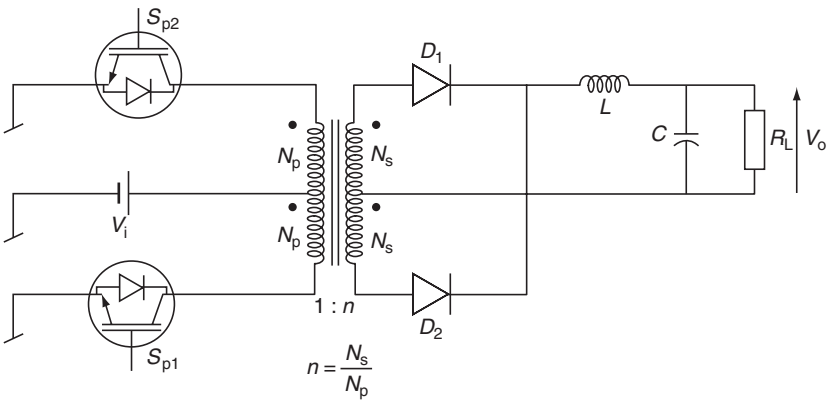


Figure 5.40 Circuit schematic of the push–pull converter with power semiconductor devices.

Secondary Side Switch Selection

The peak current requirement for the secondary side diodes, D_1 and D_2, is

$$I_{dm} > I_o + \frac{\Delta i_L}{2} \tag{5.58}$$

The average current requirement for D_1 and D_2 is

$$I_{d\text{-}av} = I_o\left(\frac{DT_s}{2T_s}\right) + \frac{I_o}{2}\left(\frac{(1-D)T_s}{T_s}\right) = \frac{I_o}{2} \tag{5.59}$$

The rms current requirement for D_1 and D_2 is

$$I_{d\text{-}rms} = \sqrt{I_o^2\left(\frac{DT_s}{2T_s}\right) + \frac{I_o^2}{4}\left(\frac{(1-D)T_s}{T_s}\right)} = I_o\sqrt{\frac{D}{2} + \frac{1-D}{4}} \tag{5.60}$$

The *PIV* requirements for the diodes D_1 and D_2 is

$$PIV > 2nV_{imax} \tag{5.61}$$

One should note that the push–pull circuit is not a robust topology in terms of circuit operation. The flux in the core swings between $-\phi_m$ and $+\phi_m$ and remains in this band only if the voltage across the primary has an average value of zero. However, in a practical circuit S_{p1} and S_{p2} do not have identical ON-state drops. Let V_{cesat1} and V_{cesat2} be the ON-state drops of S_{p1} and S_{p2}, respectively. Then the voltage across the primary when S_{p1} is ON is given by $V_i - V_{cesat1}$ and the voltage across the primary when S_{p2} is ON is determined by $V_i - V_{cesat2}$. As the pulse duration is DT_s for both the S_{p1} and S_{p2} ON periods, there will be a finite non-zero average value across the primary voltage as the positive pulse and the negative pulse amplitudes are different. This finite average voltage value across the primary will cause the flux in the core to drift towards $+\infty$ or $-\infty$ due to the relation

$$\phi = \frac{1}{N_p}\int v_p \, dt$$

This drifting flux will ultimately saturate the core. This phenomenon is called *flux walking*. To solve this problem, additional circuitry is required to sense the flux walking and appropriately adjust the duty ratio so as to reset the primary average voltage to zero. This will arrest the flux drift accordingly. However, this will increase the cost and lower the reliability of the circuit.

There is also another problem with the push–pull circuit. The coupling between the primary and the secondary is not 100%. Therefore, there will be a finite leakage inductance component as seen from the primary. The energy stored in the leakage inductance will cause large voltage spikes to occur across the OFF-state switch. The energy stored in the leakage inductances can be removed by providing alternate paths for them through circuits called the snubber circuits as discussed in Chapter 2. Due to the above drawbacks, this topology is not a very popular one.

Half-Bridge Converter

For a given power rating, the push–pull circuit requires devices that should have OFF-state voltage withstanding capability that is greater than twice the input voltage V_i. The half-bridge converter reduces the OFF-state voltage requirement of the primary side switches to V_i apart from maintaining the bi-directional flux swing in the core. Thus the voltage stress and cost of the power switches is significantly reduced as compared to the push–pull topology. The schematic of the half-bridge topology is shown in Figure 5.41.

There is no change in the secondary side circuit. It is exactly same as the secondary side circuit of the push–pull converter. On the primary side, one end of the primary winding is connected to a SPDT switch with pole P_p as shown. The other end is connected to the center of a capacitor divider as shown. The voltage at this capacitor divider point is at $V_i/2$.

The operation of the half-bridge converter circuit is illustrated in Figure 5.42. Here again, there are three operative modes like in the case of the push–pull converter circuit. The primary switches S_{p1} and S_{p2} switch during alternate cycles and their switching frequency is half the output inductor current ripple frequency.

Figure 5.42(a) shows the operative circuit when S_{p1} is ON. This occurs during the first DT_s period. The input source voltage V_i is connected to one end of the primary winding through switch S_{p1}. The other end is at $V_i/2$. The voltage across the primary is $V_i/2$ with the dot end being positive with respect to the

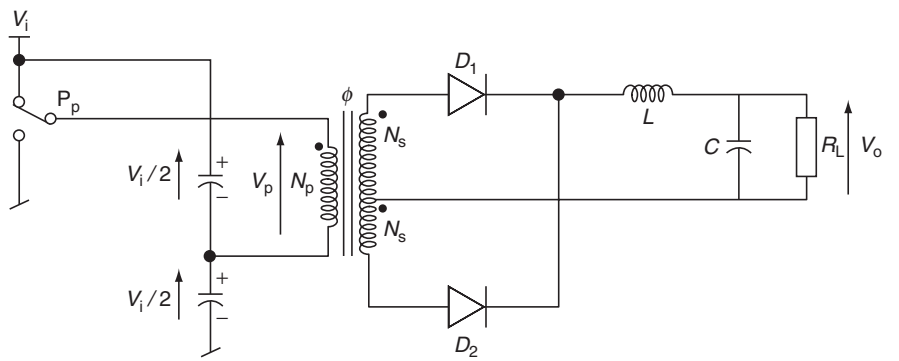


Figure 5.41 | Half-bridge converter schematic.

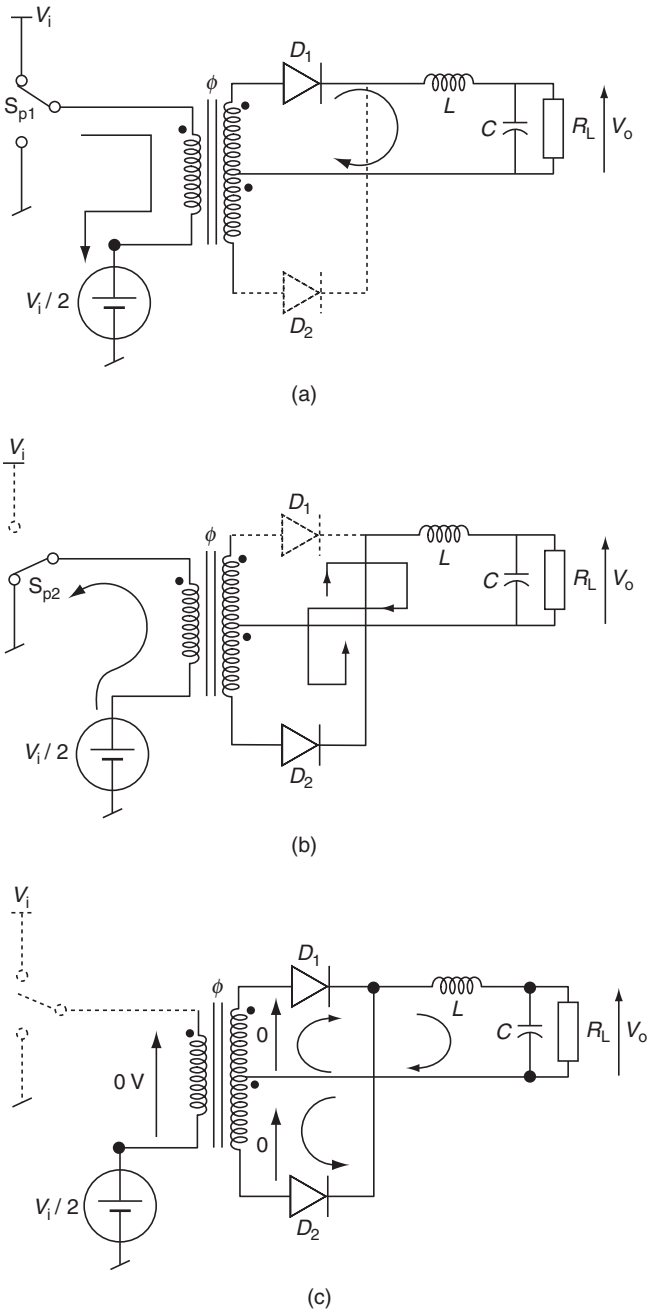


Figure 5.42 Half-bridge topology: (a) Operative circuit during DT_s (switch S_{p1} ON); (b) operative circuit during DT_s (switch S_{p2} ON); (c) operative circuit during $(1 - D)T_s$ (switches S_{p1} and S_{p2} OFF).

non-dot end. The diode D_1 is ON and D_2 is OFF. The energy is supplied to charge up the inductor through diode D_1.

Figure 5.42(b) shows the operative circuit when S_{p2} is ON. This occurs during the second DT_s period. S_{p1} and S_{p2} will be ON during alternate DT_s periods with D and T_s being defined with respect to the inductor current ripple. During this time, S_{p2} is ON and the voltage across the primary winding is $V_i/2$ but the dot end is negative with respect to the non-dot end. This will facilitate the core flux to swing in the negative direction. In the secondary, D_2 is ON and D_1 is reverse-biased. The inductor energy is replenished through diode D_2.

Figure 5.42(c) shows the operative circuit when both S_{p1} and S_{p2} are OFF. During this time, the inductor current freewheels through both D_1 and D_2. The currents flowing in the center-tapped secondary are in directions that will cancel the flux in the core due to each other. This will result in $d\phi/dt$ being zero and as a consequence the voltages across all the windings of the transformer being zero.

The circuit schematic of the half-bridge circuit with IGBT switches is shown in Figure 5.43. Switches S_{p1} and S_{p2} are replaced with IGBT switches with internal body diodes. C_1 and C_2 are two equal capacitances that form the capacitive divider. The resistances R_1 and R_2 are charge equalization resistors which will balance the charges on the capacitors C_1 and C_2 such that the center point is at $V_i/2$. One more component, the flux walking capacitor (C_{fw}) is included in series with the winding here to prevent the flux walking phenomenon. Here too, S_{p1} and S_{p2} may have non-identical ON-state drops. This will result in a finite non-zero average value across the primary winding causing the flux to walk away and saturate the core. If a capacitor C_{fw} is included in series with the winding, it will block the DC or average voltage that results due to non-identical ON-state drops. This will ensure that the primary winding will always have zero average voltage across it and hence flux walking phenomenon is absent.

Another advantage over the push–pull topology is that the body diodes are active here in removing energy stored in the leakage inductances. If S_{p1} is being turned OFF, the energy stored in the primary leakage inductance flows through the *winding–C_1–V_i–diode of S_{p2}–winding* and *winding–C_2–diode of S_{p2}–winding* paths. If S_{p2} is being turned OFF, the energy stored in the primary leakage inductance flows through the *winding–diode of S_{p1}–V_i–C_2–winding* and *winding–diode of S_{p1}–C_1–winding* paths.

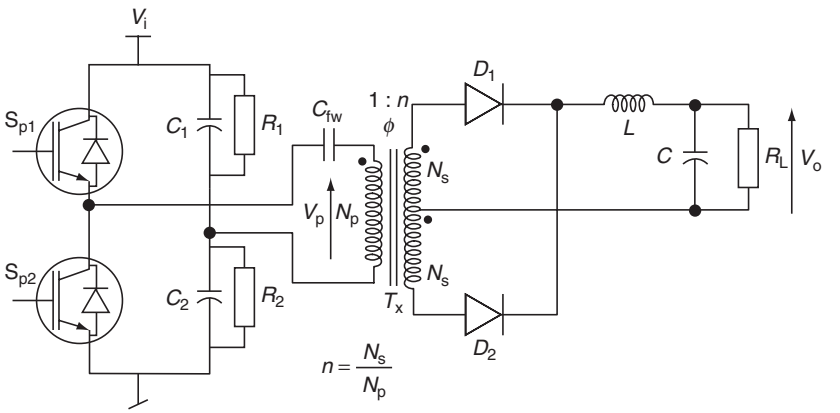


Figure 5.43 | Half-bridge circuit implementation.

The waveforms for the various signals of the half-bridge circuit are shown in Figure 5.44. The voltage across the primary is $V_i/2$ during the period when S_{p1} is ON and it is $-V_i/2$ during the time when S_{p2} is ON. The slopes for the core flux are $V_i/2N_p$ and $-V_i/2N_p$ as shown in Figure 5.44. All other waveforms are similar to that discussed for the push–pull topology.

The governing equations for the half-bridge converter are

$$V_o = n\frac{V_i}{2}D \tag{5.62}$$

where n is the secondary center tap to primary turns ratio and D is the duty ratio as defined for the buck converter inductor current waveform.

Primary-Side Switch Selection

Referring to the waveforms of Figure 5.44, the current rating for the primary side switches S_{p1} and S_{p2} is

$$I_{cm} > nI_o + \frac{n\Delta i_L}{2} \tag{5.63}$$

where I_{cm} is the peak current requirement of the IGBT. Referring to Figure 5.44 again, the peak OFF-state voltage that the primary side switches should withstand is

$$V_{ceo} > V_{imax} \tag{5.64}$$

Secondary-Side Switch Selection

The peak current requirement for the secondary side diodes, D_1 and D_2, is

$$I_{dm} > I_o + \frac{\Delta i_L}{2} \tag{5.65}$$

The average current requirement for D_1 and D_2 is

$$I_{d\text{-}av} = I_o\left(\frac{DT_s}{2T_s}\right) + \frac{I_o}{2}\left(\frac{(1-D)T_s}{T_s}\right) = \frac{I_o}{2} \tag{5.66}$$

The rms current requirement for D_1 and D_2 is

$$I_{d\text{-}rms} = \sqrt{I_o^2\left(\frac{DT_s}{2T_s}\right) + \frac{I_o^2}{4}\left(\frac{(1-D)T_s}{T_s}\right)} = I_o\sqrt{\frac{D}{2} + \frac{1-D}{4}} \tag{5.67}$$

The PIV requirements for the diodes D_1 and D_2 is

$$PIV > nV_{imax} \tag{5.68}$$

The capacitance dividers C_1 and C_2 are actually the output capacitors of the diode–capacitor filter circuit. The design and selection of C_1 and C_2 is discussed in Chapter 3. R_1 and R_2 are charge equalization resistors. An empirical rule is to choose R_1 and R_2 such that they allow around 2% of the load current to flow through them. Thus,

$$R_1 = R_2 = \frac{V_i}{2(0.02 \times I_o)}$$

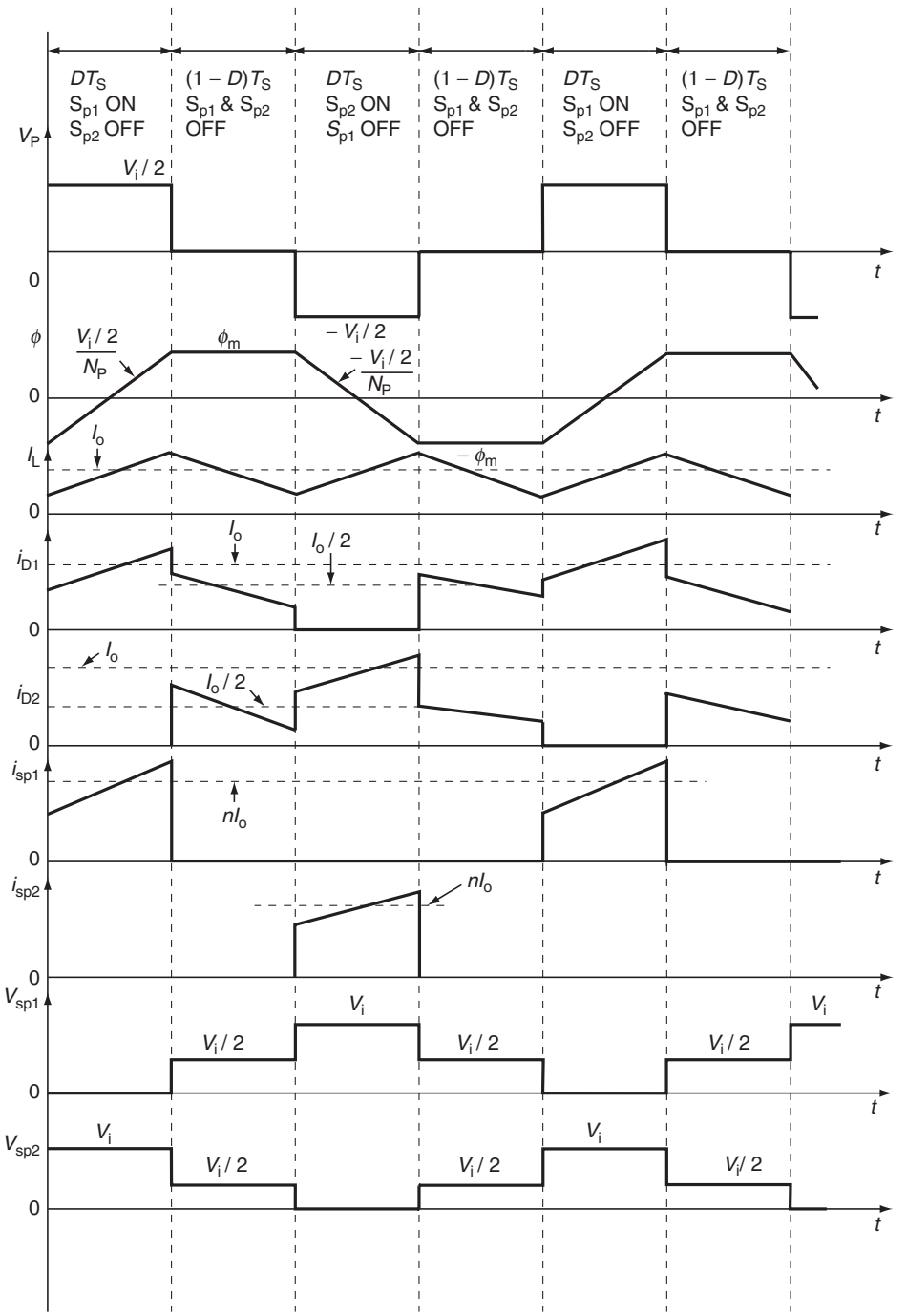


Figure 5.44 | Waveforms of the various signals of the half-bridge circuit.

Flux Walking Capacitor

Referring to the Figure 5.44, the charge added to the capacitor during the time DT_s when the switch S_{p1} is ON is

$$\Delta Q = C_{fw}\Delta V_{fw} = nI_o DT_s$$

This gives

$$C_{fw} > \frac{nI_o D}{\Delta V_{fw} f_s} \tag{5.69}$$

where D and f_s are the duty ratio and the switching frequency respectively as defined for the output side inductor current ripple. ΔV_{fw} is a design choice and it is chosen as 10% of V_i. There is another constraint that the flux walking capacitor should satisfy. The resonant frequency of the reflected secondary side inductor and the flux walking capacitor should be very much lower than the switching frequency otherwise the waveforms of the various signals will be oscillatory within a switching cycle. Thus,

$$L_p = L\left(\frac{N_p}{N_s}\right)^2$$

where L_p is the secondary inductance as seen from the primary and L is the secondary side buck converter filter inductance. The resonant frequency of the $L_p C_{fw}$ should be less than f_s by a factor of 10. Thus,

$$\frac{1}{2\pi\sqrt{L_p C_{fw}}} < \frac{f_s}{10}$$

This gives the following inequality:

$$C_{fw} > \frac{100}{4\pi^2 f_s^2 L_p} \tag{5.70}$$

Both the inequalities [Eqs. (5.69) and (5.70)] should be satisfied while selecting the capacitor value. The rms value of the current through the flux walking capacitor is given as

$$I_{Cfw\text{-}rms} = nI_o\sqrt{D} \tag{5.71}$$

Full-Bridge Converter

The half-bridge circuit applies a voltage of $V_i/2$ at the primary winding whereas a full-bridge circuit applies a voltage of V_i at the primary winding. Thus for a given power, the average current flowing through the primary of the full-bridge converter is half of that flowing through the half-bridge converter. Thus the current-carrying capacity for the full-bridge converter switches is less than that of the switches of the half-bridge converter of the same output power rating. All other operational features of the full-bridge converter are same as those of the half-bridge converter. The schematic of the full-bridge topology is shown in Figure 5.45.

There is no change in the secondary side circuit. It is exactly same as the secondary side circuit of the half-bridge converter. On the primary side, one end of the primary winding is connected to an SPDT switch with pole P_{p1} as shown. The other end is connected to another SPDT switch with pole P_{p2} as shown.

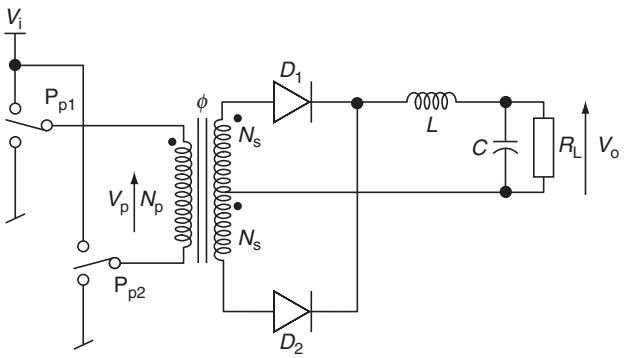


Figure 5.45 | Full-bridge converter topology.

The operation of the full-bridge converter circuit is illustrated in Figure 5.46. Here again, there are three operative modes like in the case of the push–pull and the half-bridge converter circuits. The primary switches S_{p1} and S_{p4} are linked and they switch ON and OFF simultaneously. Likewise the primary switches S_{p2} and S_{p3} are linked and they switch ON and OFF simultaneously. The pairs $S_{p1,4}$ and $S_{p2,3}$ switch ON during alternate cycles and their switching frequency is half the output inductor current ripple frequency as discussed for the push–pull and half-bridge converter circuits.

Figure 5.46(a) shows the operative circuit when $S_{p1,4}$ are ON. This occurs during the first DT_s period. The input source voltage V_i is connected to one end of the primary winding through switch S_{p1}. The other end is connected to ground through S_{p4}. The voltage across the primary is V_i with the dot end being positive with respect to the non-dot end. The diode D_1 is ON and D_2 is OFF. The energy is supplied to charge up the inductor through diode D_1.

Figure 5.46(b) shows the operative circuit when $S_{p2,3}$ are ON. This occurs during the second DT_s period. During this time, S_{p2} and S_{p3} are ON and the voltage across the primary winding is V_i but the dot end is negative with respect to the non-dot end. This will facilitate the core flux to swing in the negative direction. In the secondary, D_2 is ON and D_1 is reverse-biased. The inductor energy is replenished through diode D_2.

Figure 5.46(c) shows the operative circuit when $S_{p1,4}$ and $S_{p2,3}$ are OFF. During this time, the inductor current freewheels through both D_1 and D_2. The currents flowing in the center-tapped secondary are in directions that will cancel the flux in the core due to each other. This will result in $d\phi/dt$ being zero and as a consequence the voltages across all the windings of the transformer are zero.

The circuit schematic of the full-bridge circuit with IGBT switches is shown in Figure 5.47. The switches S_{p1} to S_{p4} are replaced with IGBT switches with internal body diodes. C_{fw} is included in series with the winding here. This is called the flux walking capacitor that will prevent the flux walking phenomenon in a manner similar to that discussed for the half-bridge circuit.

Here the body diodes are active in removing energy stored in the leakage inductances as seen from the primary. If $S_{p1,4}$ is being turned OFF, the energy stored in the primary leakage inductance flows through the *winding–S_{p3} diode–V_i–S_{p2} diode–winding* path. If $S_{p2,3}$ is being turned OFF, the energy stored in the primary leakage inductance flows through the *winding–S_{p1} diode–V_i–S_{p4} diode–winding* path.

The waveforms for the various signals of the full-bridge circuit are shown in Figure 5.48. The voltage across the primary is V_i during the period when $S_{p1,4}$ are ON and it is $-V_i$ during the time when $S_{p2,3}$

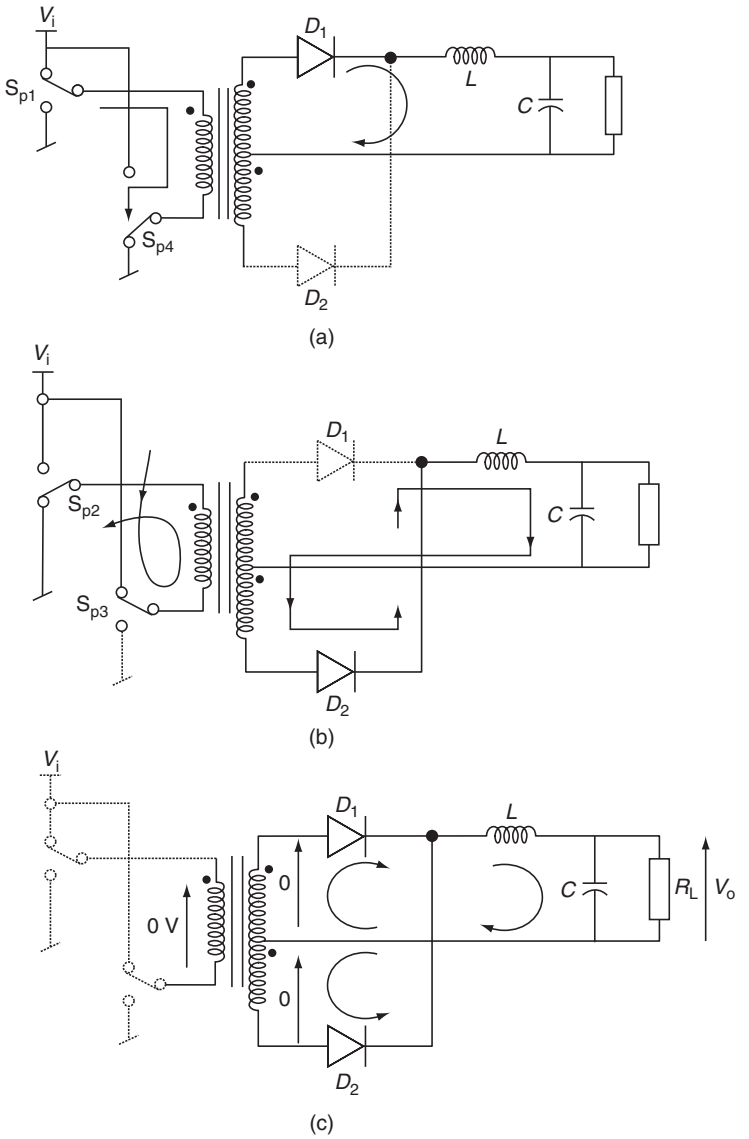


Figure 5.46 Full-bridge topology: (a) Operative circuit during DT_s (switch $S_{p1,4}$ ON); (b) operative circuit during DT_s (switch $S_{p2,3}$ ON); (c) operative circuit during $(1 - D)T_s$ (switches $S_{p1,4}$ and $S_{p2,3}$ are OFF).

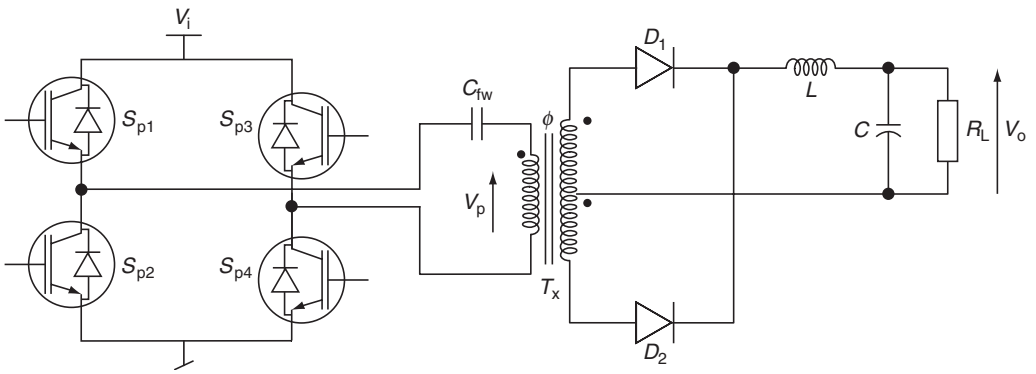


Figure 5.47 | Full-bridge circuit implementation.

are ON. The slopes for the core flux are V_i/N_p and $-V_i/N_p$ as shown in Figure 5.48. All other waveforms are similar to that discussed for the half-bridge topology.
The governing equations for the full-bridge converter are

$$V_o = nV_iD \tag{5.72}$$

where n is the secondary center tap to primary turns ratio and D is the duty ratio as defined for the buck converter inductor current waveform.

Primary Side Switch Selection

Referring to the waveforms of Figure 5.48, the current rating for the primary side switches S_{p1} to S_{p4} is

$$I_{cm} > nI_o + \frac{n\Delta i_L}{2} \tag{5.73}$$

where I_{cm} is the peak current requirement of the IGBT. Referring to Figure 5.48 again, the peak OFF-state voltage that the primary side switches should withstand is

$$V_{ceo} > V_{imax} \tag{5.74}$$

Secondary Side Switch Selection

The peak current requirement for the secondary side diodes, D_1 and D_2, is

$$I_{dm} > I_o + \frac{\Delta i_L}{2} \tag{5.75}$$

The average current requirement for D_1 and D_2 is

$$I_{d\text{-}av} = I_o\left(\frac{DT_s}{2T_s}\right) + \frac{I_o}{2}\left(\frac{(1-D)T_s}{T_s}\right) = \frac{I_o}{2} \tag{5.76}$$

The rms current requirement for D_1 and D_2 is

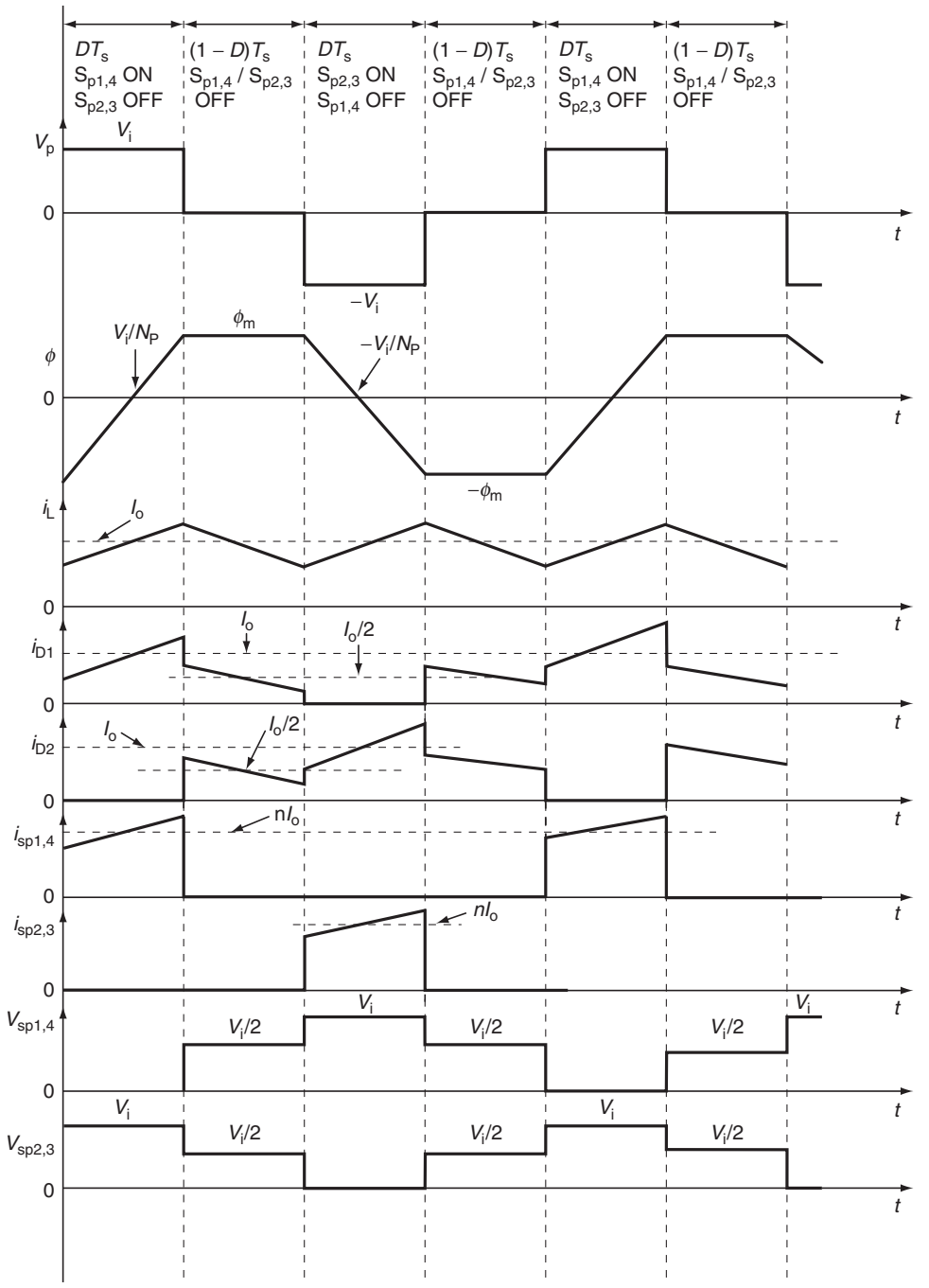


Figure 5.48 | Waveforms for the various signals of the full-bridge converter.

$$I_{\text{d-rms}} = \sqrt{I_{\text{o}}^2\left(\frac{DT_{\text{s}}}{2T_{\text{s}}}\right) + \frac{I_{\text{o}}^2}{4}\left(\frac{(1-D)T_{\text{s}}}{T_{\text{s}}}\right)} = I_{\text{o}}\sqrt{\frac{D}{2} + \frac{1-D}{4}} \tag{5.77}$$

The *PIV* requirement for the diodes D_1 and D_2 is

$$PIV > 2nV_{\text{imax}} \tag{5.78}$$

Flux Walking Capacitor

The design of the flux walking capacitor is same as that discussed for the half-bridge converter circuit. The capacitance value of the flux walking capacitor should be chosen based on

$$C_{\text{fw}} > \frac{nI_{\text{o}}D}{\Delta V_{\text{fw}} f_{\text{s}}} \tag{5.79}$$

where D and f_{s} are the duty ratio and the switching frequency, respectively, as defined for the output side inductor current ripple. ΔV_{fw} is a design choice and it is chosen as 10% of V_{i}.

$$C_{\text{fw}} > \frac{100}{4\pi^2 f_{\text{s}}^2 L_{\text{p}}} \tag{5.80}$$

where L_{p} is the secondary inductance as seen from the primary given by

$$L_{\text{p}} = L\left(\frac{N_{\text{p}}}{N_{\text{s}}}\right)^2$$

Both the inequalities [i.e., Eqs. (5.79) and (5.80)] should be satisfied while selecting the capacitor value. The rms value of the current through the flux walking capacitor is given by

$$I_{\text{Cfw-rms}} = nI_{\text{o}}\sqrt{D} \tag{5.81}$$

Flyback Converter

Flyback converter is an isolated converter that is based on the primary buck–boost converter topology. This converter has the least component count and hence is a very popular topology. Figure 5.49 shows the development of the flyback converter topology. Figure 5.49(a) shows the primary buck–boost topology that is discussed earlier section. Figure 5.49(b) shows the same converter with the SPDT switch shifted to the ground side. The inductor is designed like in the non-isolated buck–boost converter case. However, one more winding called the secondary is wound on top of the existing inductor winding to provide galvanic isolation. The SPDT switch is split into two SPST switches, each being connected to one winding of the two winding inductor. This is shown in Figure 5.49(c). Figure 5.49(d) shows the secondary side modified so that the output voltage is measured in the conventional positive sense. Observe the dot polarities on the windings implying the out of phase switching of S_{p1} and S_{s1}.

Figure 5.50 illustrates the operation of the flyback converter. There are two operating states for the converter. One is the period when the inductor energy builds up or the DT_{s} period and the other is the period when the inductor energy discharges to the load or the $(1 - D)T_{\text{s}}$ period. Figure 5.50(a) shows the operative circuit during the DT_{s} period. During this time, the switch S_{p1} is ON and the switch S_{s1} in the secondary side is OFF. The dot ends of the windings are positive. The inductor energy builds up as it is now directly

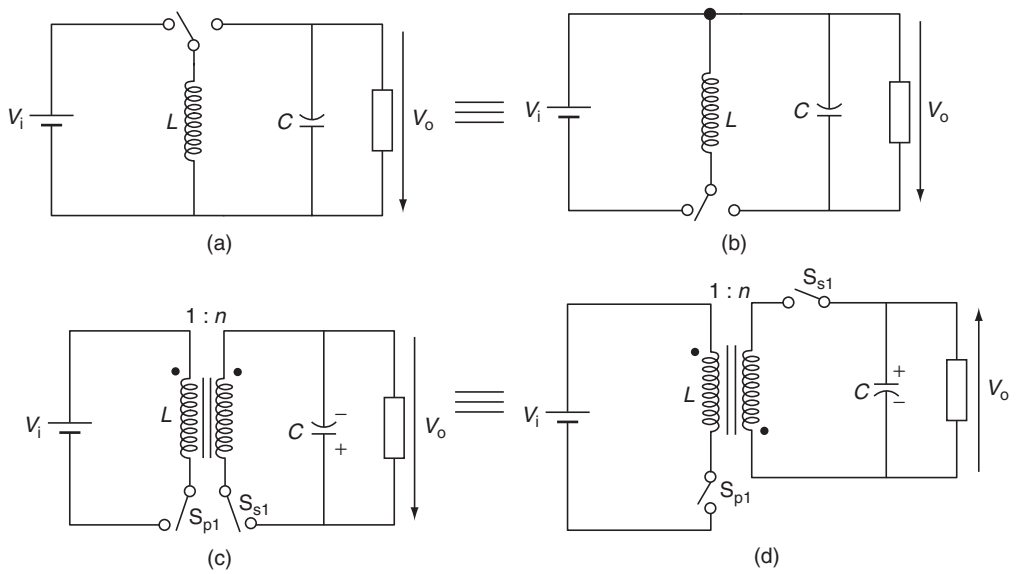


Figure 5.49 Development of the flyback converter topology from the primary buck–boost converter.

connected across the input source V_i. At this same time, the capacitor discharges into the load. During this period there is transfer of energy from the primary to the secondary.

Figure 5.50(b) shows the operative circuit during the $(1 - D)T_s$ period. During this time, the switch S_{p1} is OFF and the switch S_{s1} is ON. The winding voltages reverse polarity and the non-dot ends become positive with respect to the dot ends. The primary winding of the inductor is out of action. The secondary winding now discharges the energy stored in the core to the capacitor and the load.

The SPST switches are replaced by power semiconductor switches in Figure 5.51. The switch S_{p1} is replaced by an IGBT and the switch S_{s1} is replaced by the diode depending on the *i*–*v* characteristic requirements of the switches. The waveforms at various points on the flyback converter circuit are shown in Figure 5.52.

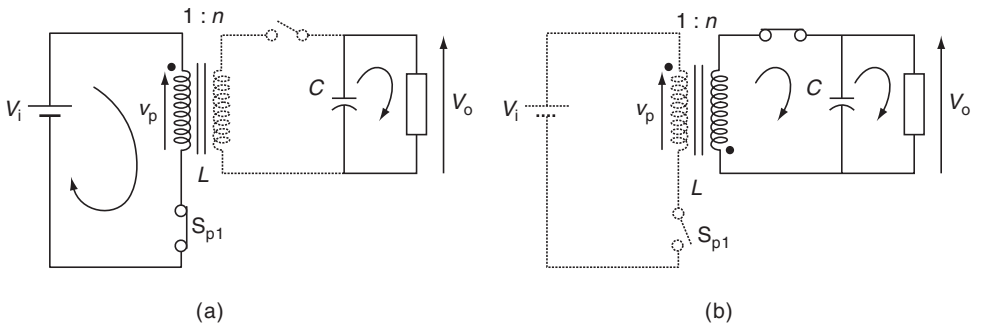


Figure 5.50 Flyback converter: (a) Operative circuit during DT_s period; (b) operative circuit during $(1 - D)T_s$ period.

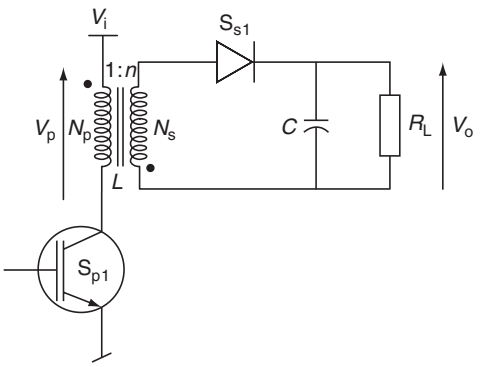


Figure 5.51 | Schematic of the flyback topology implementation.

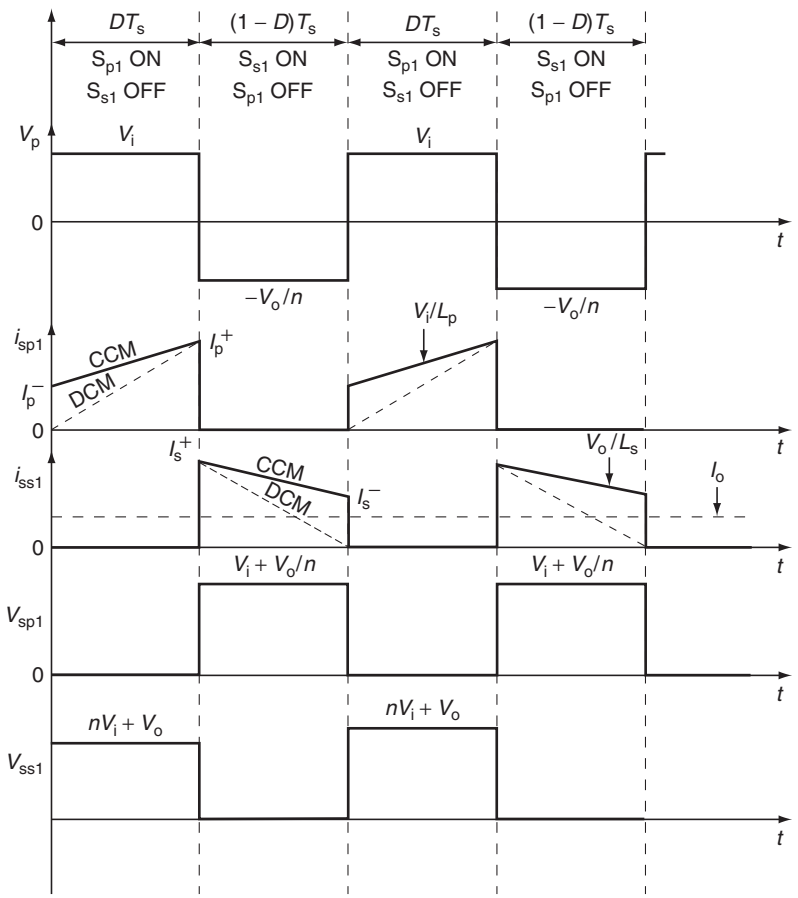


Figure 5.52 | Waveforms of various signals of the flyback converter.

During the DT_s period, the switch S_{p1} is ON. The voltage across the primary is V_i with the dot end positive. The current through the primary inductor and the switch i_{sp1} increases with a rate of V_i/L_p where L_p is the inductance as seen from the primary. The voltage across the secondary is nV_i with the dot end positive. During this time the diode switch S_{s1} is OFF. The voltage across S_{s1} is $nV_i + V_o$.

During the $(1 - D)T_s$ period, the switch S_{s1} is ON. The voltage across the secondary is V_o and the voltage across the primary is V_o/n with the dot end negative. The stored inductor energy now flows through the secondary diode and into the capacitor and the load. The rate of this current is V_o/L_s where L_s is the inductance as seen from the secondary. During this time the primary side IGBT switch S_{p1} is OFF. The voltage across S_{p1} is $V_i + (V_o / n)$.

Governing Equations

The design equations for selecting the switches of the flyback converter are as follows:

$$V_o = nV_i\left(\frac{D}{1-D}\right) \tag{5.82}$$

where n is the secondary to primary turns ratio and D is the duty ratio. Referring to the waveforms of Figure 5.52,

$$I_o = \left(\frac{I_{s-} + I_{s+}}{2}\right)(1-D) \tag{5.83}$$

The stored inductor energy transferred to the capacitor during $(1 - D)T_s$ period has to supply power P_o to the load during $(1 - D)T_s$ period and also to replenish the energy lost by the capacitor during the previous DT_s period. Thus, P_oT_s amount of energy has to supplied from the stored inductor energy to the load, during the $(1 - D)T_s$ period. Thus,

$$P_oT_s = \frac{1}{2}L_s(I_{s+}^2 - I_{s-}^2) = \frac{1}{2}L_s(I_{s+} + I_{s-})(I_{s+} - I_{s-}) \tag{5.84}$$

From Eqs. (5.83) and (5.84), one obtains

$$I_{s+} - I_{s-} = \frac{V_o(1-D)}{L_sf_s} \tag{5.85a}$$

From Eqs. (5.83) and (5.85a), one obtains

$$I_{s+} = \frac{I_o}{1-D} + \frac{V_o(1-D)}{2L_sf_s} \tag{5.85b}$$

$$I_{s-} = \frac{I_o}{1-D} - \frac{V_o(1-D)}{2L_sf_s} \tag{5.85c}$$

Similarly on the primary side,

$$I_{in} = \frac{P_i}{V_i} = \left(\frac{I_{p+} + I_{p-}}{2}\right)D \tag{5.86}$$

Assuming 100% efficiency, the energy that is put into the inductor during DT_s period is

$$P_o \cdot T_s = \frac{1}{2} L_p (I_{p+}^2 - I_{p-}^2) = \frac{1}{2} L_p (I_{p+} + I_{p-})(I_{p+} - I_{p-}) \tag{5.87}$$

From Eqs. (5.86) and (5.87), one obtains

$$I_{p+} - I_{p-} = \frac{V_i D}{L_p f_s} \tag{5.88}$$

From Eqs. (5.86) and (5.88), one obtains

$$I_{p+} = \frac{I_{in}}{D} + \frac{V_i D}{2 L_p f_s} \tag{5.89}$$

$$I_{p-} = \frac{I_{in}}{D} - \frac{V_i D}{2 L_p f_s} \tag{5.90}$$

I_{in} and V_i can be obtained from the specified I_o and V_o values using the input–output relationship for the flyback converter topology as given in Eq. (5.82).

For DCM operation, I_{p-} and I_{s-} will be zero. In DCM operation, the switching period will be divided into three parts: (a) $D_1 T_s$ duration when the primary switch (IGBT) is ON, (b) $D_2 T_s$ duration when the secondary diode is ON and (c) the remaining duration when both the switches are OFF. The current requirement for the S_{p1} IGBT switch is given as

$$I_{cm} > I_{p+} \tag{5.91}$$

where I_{cm} is the peak current requirement of the IGBT. The current requirement for the S_{s1} diode switch is given as

$$I_{dm} > I_{s+} \tag{5.92}$$

The average current requirement for the S_{s1} diode switch is

$$I_{d\text{-}avg} = I_o \tag{5.93}$$

The rms current requirement for the S_{s1} diode switch is

$$I_{d\text{-}rms} = \left(\frac{I_{s-} + I_{s+}}{2} \right) \sqrt{1 - D} \tag{5.94}$$

Referring to Figure 5.52 again, the peak OFF-state voltage that the primary side switch S_{p1} should withstand is

$$V_{ceo} > V_{imax} + \frac{V_o}{n} \tag{5.95}$$

The peak OFF-state voltage that the secondary side switch S_{s1} should withstand is

$$PIV > nV_i + V_o \tag{5.96}$$

5.8 Practical Issues

In the previous sections the three primary converters and the isolated converters were discussed. However, there are a few practical issues that are not directly evident from the steady-state analysis. One should be wary of the following problems when actually implementing the converter circuits:

1. effects due to leakage inductance, track or wire inductance and stray series inductance;
2. effects due to parasitic capacitances across the switching device;
3. current shoot-through during device turn-ON.

All practical transformers and magnetic couplings will always have a finite non-zero leakage inductance as seen from a given winding. The voltage drops across these inductances that result due to the switch OFF of the current through it will invariably result in large voltage spikes across the switch being switched OFF. If either the leakage inductance is high due to poor magnetic coupling between windings or the di/dt for the current through the switch which is being cut-off is high, then the resulting voltage spike that occurs across the switch will lead to deteriorated operation of the switch due to increased voltage stress if not damaging the device. Therefore, the issue of leakage inductance is a very important issue. There are, in general, two solutions that are commonly adopted: (a) reduce the leakage inductance by using tightly wound windings and (b) divert the energy stored in the leakage inductance. In the former solution, close coupling between windings will increase the inter-winding capacitance and thereby reduce the differential voltage capacity between the windings as high differential can lead to arcing or break over of the insulation. The latter solution is addressed by means of snubber circuits. These circuits are discussed in Chapter 2.

The power semiconductor switches have significant capacitance across their collector–emitter or drain–source terminals. When the switch is OFF, the capacitance across the device will charge up to the OFF-state voltage. Again when the switch is turned ON, the charged up capacitance sees a very low impedance path through the ON-state device and discharges through it limited only by the ON-state resistance. Thus a very large current spike will flow through the device at turn-ON that may damage the device. This is solved by using series snubber circuits. Alternately, this can be addressed by using zero-current switching (ZCS) or zero-voltage switching (ZVS) topologies. This will be discussed in a later section.

In many topologies, for example the half-bridge and full-bridge converters, there is a serious problem of current shoot-through. In the case of a bridge arm, when the top device of a bridge arm is ON, the corresponding bottom device is OFF. On switching OFF the top device and turning ON the bottom device, the top device would not have recovered completely to the OFF-state as the turn-OFF process is slower than the turn-ON process. Thus, there is a very small but finite period of time during the switching transition when the input voltage source V_i is directly connected to the ground which will result in a very large current spike. This will eventually damage the devices. There are again two solutions that are commonly adopted: (a) providing dead time between the top and bottom devices to give enough time for the switching OFF device to turn OFF and (b) using snubber circuits that are discussed in Chapter 2. It should be noted that the above problem is not specific to only half-bridge or full-bridge circuits. It occurs in many topologies of the converters discussed previously, in both primary converters and isolated converters.

5.9 Other Topologies

The topologies discussed till now are the more common and standard topologies that are very popular. However, there are numerous topologies that have been adapted from the primary converters for various applications. In fact using the rules stated at the start of this chapter, one may construct different

topologies such that the volt-second balance, charge balance, flux reset and *i–v* characteristic requirements for the switches are ensured. All the topologies can be analyzed in a similar manner as presented in the previous sections. The steady-state input–output relationship, the inductor and capacitor values, the switch ratings and the type of semiconductor switches can be deduced from the waveforms of the various signals of the converter. In this section, few topologies like the Cuk converter, the high-boost converter, *zero-current switching* (*ZCS*) and *zero-voltage switching* (*ZVS*) converters will be discussed briefly to give a flavor of their operation and application.

Cuk Converter

The Cuk converter is named after the inventor of the converter. It consists of two inductors, one placed at the input side like the boost topology and the other placed at the output side like the buck topology. The topology also has two capacitors C_1 and C_2 as shown in Figure 5.53. There is an SPDT switch placed such that the pole is connected to the circuit ground and the two throws T1 and T2 are connected, as shown, across the capacitor C_1. The output V_o that is obtained from this topology is negative with respect to the circuit ground as will become clear from the circuit operation. It is interesting to note that both the input side and the output side contain an inductor each. This implies that the input and output currents of this converter are not switching or pulsed currents. Figure 5.54 shows the operative modes of the Cuk converter topology.

Referring to Figure 5.54, there are two operative modes for the Cuk converter circuit. During the DT_s period as shown in Figure 5.54(a), the SPDT switch is connected to throw T1 (i.e., S1 is ON and S2 is OFF). The energy in the inductor L_1 builds up by the current flowing from the input source V_i. The energy in the inductor L_2 also builds up due to transfer of the stored energy from capacitor C_1 to the inductor and capacitor C_2. Observe the direction of flow of the L_2 inductor current. It charges the output capacitor C_2 in such a direction that V_o is negative with respect to the circuit ground.

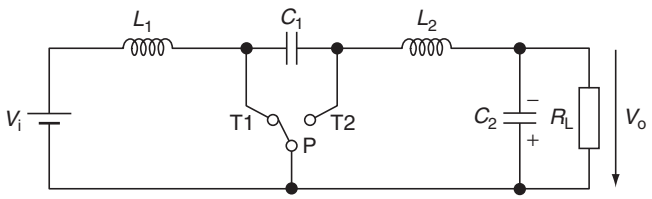


Figure 5.53 | Schematic of the Cuk converter topology.

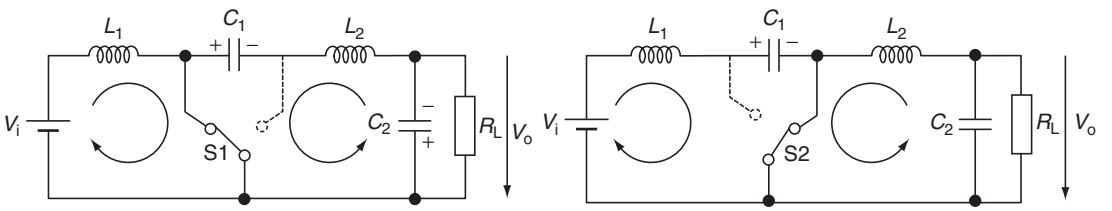


Figure 5.54 | Cuk converter: (a) Operative circuit during DT_s period; (b) operative circuit during $(1 - D)T_s$ period.

During the $(1 - D)T_s$ period, the operative circuit is as shown in Figure 5.54(b). During this time, the pole of the SPDT switch is connected to the throw T2 (i.e., S2 is ON and S1 is OFF). The energy that was stored in the inductor L_1 during the DT_s period is now transferred to the capacitor C_1. The inductor current in L_2 freewheels through the S2 switch just like in the buck converter case. The energy of inductor L_2 is now transferred to the capacitor C_2 and the load R_L.

Input–Output Relationship

To obtain the input–output voltage relationship, apply the volt-second balance rule to both the inductors. This implies that the area under the inductor voltage waveforms in one period under steady-state conditions should be zero.

Volt-Second Balance for L_1 During DT_s period the voltage across the inductor L_1 is V_i and during $(1 - D)T_s$ period, the voltage across the inductor is $V_i - V_{C1}$. Thus,

$$(V_i)DT_s + (V_i - V_{C1})(1 - D)T_s = 0 \tag{5.97}$$

where V_{C1} is the voltage across capacitor C_1. Re-arranging Eq. (5.97), one obtains

$$V_{C1} = \frac{V_i}{1 - D} \tag{5.98}$$

Volt-Second Balance for L_2 During DT_s period the voltage across the inductor L_2 is $V_{C1} - V_o$ and during $(1 - D)T_s$ period, the voltage across the inductor is $-V_o$. Thus,

$$(V_{C1} - V_o)DT_s + (-V_o)(1 - D)T_s = 0 \tag{5.99}$$

Re-arranging Eq. (5.99), one obtains

$$V_{C1} = \frac{V_o}{D} \tag{5.100}$$

From Eqs. (5.98) and (5.100), the voltage input–output relationship is given as

$$V_o = \frac{V_i D}{1 - D} \tag{5.101}$$

Observe that the input–output voltage relationship is similar to the buck–boost input–output voltage relationship. Figure 5.55 shows the circuit implementation schematic. Switch S1 is replaced by an IGBT and switch S2 is replaced by a diode after considering the *i–v* characteristic requirements of the two switches.

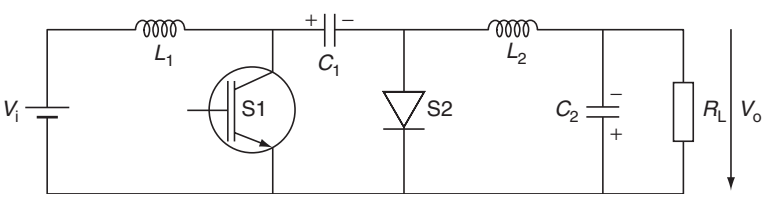


Figure 5.55 Schematic of Cuk converter circuit implementation.

Zero-Current Ripple

The Cuk converter can be configured to give additional benefits by coupling the inductors L_1 and L_2. The input side and output side inductors carry non-switching DC currents. However these currents have a ripple current riding over them. By coupling the two inductors, one may obtain either the input side or the output side inductor current with zero ripples. For example, if the output side inductor current ripple is made zero, the size of the capacitor C_2 which is dependent on the inductor ripple current will be drastically reduced if not eliminated.

The inductive coupling may be done for any converter having multiple inductors provided the instantaneous voltages across the two inductors are the same or have the same scale factors. In the case of the Cuk converter, consider the voltage across the two inductors at every instant. First, an estimate of the voltage across the capacitor C_1 in the steady state can be obtained from Eqs. (5.98) and (5.100) as

$$V_{C1} = V_i + V_o \tag{5.102}$$

This is also evident from applying the Kirchhoff's voltage law to the outer loop of the Cuk converter in the steady state when the drops across the inductors can be considered zero. Let V_{L1} and V_{L2} be the voltages across the inductors L_1 and L_2, respectively.

During DT_s period

$$V_{L1} = V_i$$

$$V_{L2} = V_{C1} - V_o = V_i$$

During $(1 - D)T_s$ period

$$V_{L1} = V_i - V_{C1} = V_i - V_i - V_o = -V_o$$

$$V_{L2} = -V_o$$

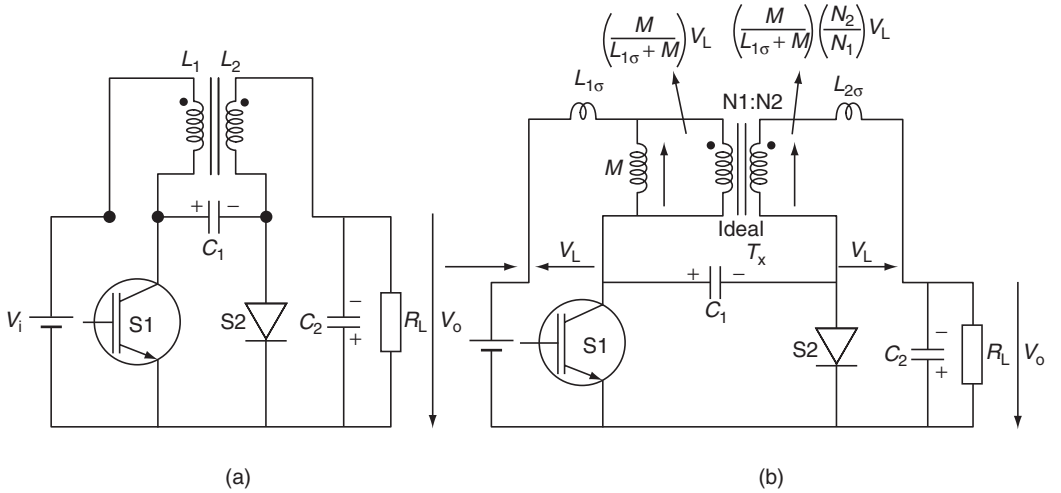


Figure 5.56 | (a) Cuk converter with inductive coupling; (b) non-idealities in the inductive coupling.

From the above it is evident that the voltage across L_1 and L_2 are same at every instant. Thus there is a case for coupling the two. Figure 5.56(a) gives the schematic of the Cuk converter with the two inductors coupled with the specified dot polarity. If the output current ripple is to be reduced to zero, the equivalent circuit of the coupled inductor portion is as shown in Figure 5.56(b). The non-idealities like the leakage inductances are put to good use in bringing down the current ripple to zero. The leakage inductance $L_{2\sigma}$ is the component in focus. If the voltage across the leakage inductance $L_{2\sigma}$ is made zero, then as $L_{2\sigma}$ is non-zero, $\mathrm{d}i/\mathrm{d}t$ through the inductor $L_{2\sigma}$ has to be zero. This implies that the ripple in the current flowing through $L_{2\sigma}$ is zero. As the current flowing through $L_{2\sigma}$ is the same as the current flowing through L_2, the L_2 current ripple is eliminated and shifted to the L_1 side. This means that C_2 can be very small or eliminated.

Referring to Figure 5.56(b), the voltage that gets translated from the L_1 side to the L_2 side is

$$\left(\frac{M}{L_{1\sigma}+M}\right)\left(\frac{N_2}{N_1}\right)V_{\mathrm{L1}}$$

The voltage across $L_{2\sigma}$ should be zero. The condition for that is

$$\left(\frac{M}{L_{1\sigma}+M}\right)\left(\frac{N_2}{N_1}\right)V_{\mathrm{L1}}-V_{\mathrm{L2}}=0$$

As $V_{\mathrm{L1}} = V_{\mathrm{L2}}$ at every instant,

$$\left(\frac{M}{L_{1\sigma}+M}\right)\left(\frac{N_2}{N_1}\right)=1$$

$$\frac{M}{L_{1\sigma}+M}=\frac{N_1}{N_2} \tag{5.103}$$

The constraint of Eq. (5.103) should be satisfied to obtain zero ripple in the L_2 inductor current. Similarly, analysis can be done to obtain zero ripple current through L_1 inductor at the input side by making the voltage across the $L_{1\sigma}$ leakage inductance zero. It should be noted that zero ripple current can be obtained either in the L_1 inductor current or the L_2 inductor current and not both together.

Isolated Cuk Converter

Isolation can be incorporated in the Cuk converter also. Figure 5.57 shows the schematic of the isolated Cuk converter. The capacitor C_1 is split into two equal parts as $C_{1\mathrm{p}}$ and $C_{1\mathrm{s}}$ and the transformer is placed in between them as shown. The input–output relationship can be derived as discussed before for other converters by using the volt-second balance. Thus,

$$V_{\mathrm{o}}=\frac{nV_{\mathrm{i}}D}{1-D} \tag{5.104}$$

where n is the secondary to primary turns ratio.

In the case of the isolated Cuk converter shown in Figure 5.57, consider the voltage across the two inductors and the two windings of the transformer at every instant. Applying the Kirchhoff's voltage law to the primary loop and the secondary loop of the Cuk converter in the steady state, it is evident that the voltage across $C_{1\mathrm{p}}$ is V_{i} and the voltage across $C_{1\mathrm{s}}$ is V_{o} as indicated in the Figure 5.57. Let V_{L1} and V_{L2} be

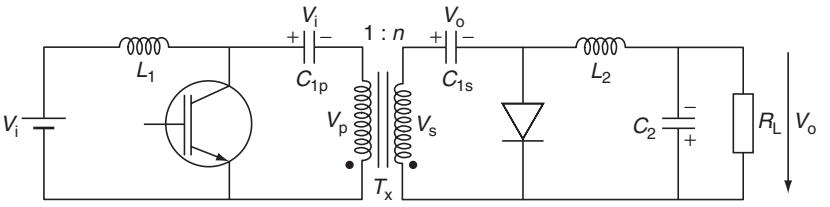


Figure 5.57 | Isolated Cuk converter.

the voltages across the inductors L_1 and L_2, respectively, and V_p and V_s be the voltages across the primary and secondary windings of the transformer, respectively.

During DT_s period

$$V_{L1} = V_i$$

$$V_p = V_i$$

$$V_s = nV_i$$

$$V_{L2} = V_{C1s} + V_s - V_o = V_o + nV_i - V_o = nV_i$$

During $(1 - D)T_s$ period

$$V_{L1} = V_i - V_{C1p} + V_p = -\frac{V_o}{n}$$

$$V_p = -\frac{V_o}{n}$$

$$V_s = -V_o$$

$$V_{L2} = -V_o$$

Thus from above one can observe that if $n = 1$, then all the windings, that is, the L_1 inductor, the L_2 inductor, the primary winding and the secondary winding, all have the same voltage across them at every instant. For n other than 1, the voltages across the primary side windings, that is L_1 and primary winding, is different from the voltages across the secondary side windings, that is L_2 and secondary winding, by a constant factor of n. Thus, there is a case to couple all the windings onto the same magnetic core. This coupled magnetic Cuk converter wherein the L_1 inductor, the L_2 inductor and the transformer windings are all wound on the same common core is shown in Figure 5.58. Here both the input side and the output side current ripples can be made zero simultaneously by transferring the ripples to the transformer. Analysis for this can be performed along similar lines as discussed for the non-isolated Cuk converter with coupled inductors.

High-Boost Converter

The boost converter discussed as one of the primary converters cannot give very high gain as explained earlier. As the duty ratio D becomes unity, the output does not become infinite but zero. In fact, a practical boost converter can give a gain of only about 4 or 5. To get a very high boost, the circuit of Figure 5.59 is a very useful and interesting topology.

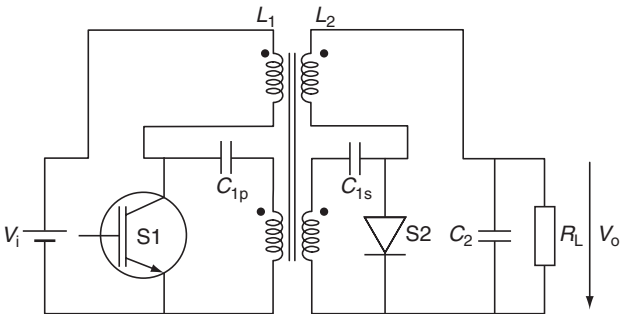


Figure 5.58 | Coupled magnetic Cuk converter.

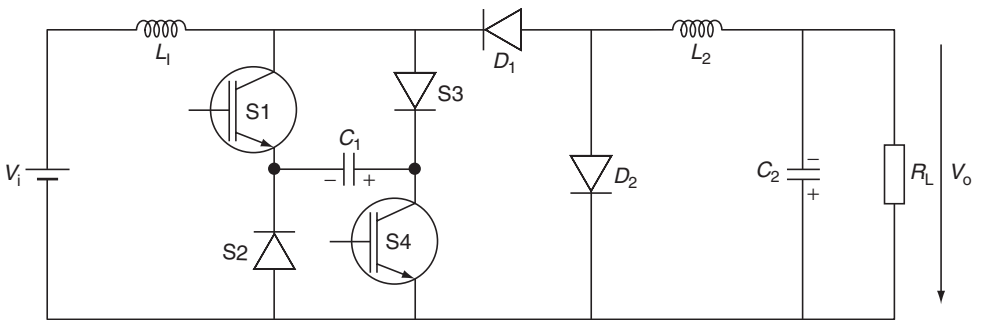


Figure 5.59 | Schematic of a high-boost converter.

This topology also consists of two inductors L_1 and L_2 as shown and four switches in a bridge configuration. During the time when S1 and S4 are OFF, the capacitor C_1 charges up to V_i with the polarity, as shown, by charging from the input source V_i through the diodes S3 and S2. When switches S1 and S4 are turned ON, the voltage across C_1 appears across the diodes S3 and S2 to reverse bias them thereby turning OFF the diodes S3 and S2. The output portion of the circuit is similar to the buck converter topology. The operation of the converter is described as follows,

During the Time DT_s

Here S1 and S4 are ON and S3 and S2 are OFF. The voltage across L_1 is given as

$$V_{L1} = V_i + V_{C1} \tag{5.105}$$

The inductor energy builds up during this time. During this time, the capacitor C_1 also appears across the output circuit in such a manner as to reverse bias diode D_2 and forward bias D_1. The capacitor C_1 energy is transferred to the output inductor wherein there is build up of energy. The voltage across L_2 is given as

$$V_{L2} = V_{C1} - V_o \tag{5.106}$$

During the Time (1 – D)T_s

In this case S1 and S4 are turned OFF and S3 and S2 are ON. The energy in the inductor L_1 charges up the capacitor C_1 through the diodes S3 and S2. The voltage across L_1 is given as

$$V_{L1} = V_i - V_{C1} \tag{5.107}$$

During this time, the current in L_2 will freewheel through D_2 transferring energy from L_2 to C_2. The diode D_1 is reverse-biased during this time and is therefore OFF. The voltage across L_2 is given as

$$V_{L2} = -V_o \tag{5.108}$$

Input–Output Relationship

The input–output relationship can be found out as before by applying the volt-second balance rule for the inductors during the steady-state operating conditions. Using Eqs. (5.105) and (5.107) and applying the volt-second balance rule for L_1, one obtains

$$(V_i + V_{C1})DT_s + (V_i - V_{C1})(1-D)T_s = 0$$

$$V_{C1} = \frac{V_i}{1-2D} \tag{5.109}$$

Using Eqs. (5.106) and (5.108) and applying the volt-second balance rule for L_2, one obtains

$$(V_{C1} - V_o)DT_s + (-V_o)(1-D)T_s = 0$$

$$V_o = V_{C1}D \tag{5.110}$$

From Eqs. (5.109) and (5.110), the following input–output relationship results:

$$V_o = \frac{V_i D}{1-2D} \tag{5.111}$$

From the above input–output relationship, it can be observed that when $D = 0.5$, V_o tends towards infinity. It can be noted that when $D = 0.5$, the converter switches are switching with 50% ON and 50% OFF durations and the steady-state input–output relationship of Eq. (5.110) is still valid. Thus one can obtain high boost or gains from this topology.

From Eqs. (5.105)–(5.109) and Eq. (5.111), it can be deduced that during the period DT_s

$$V_{L1} = \frac{2V_o(1-D)}{D}; \quad V_{L2} = \frac{V_o(1-D)}{D}$$

Likewise, during the period $(1 - D)T_s$,

$$V_{L1} = -2V_o; \quad V_{L2} = -V_o$$

Thus it can be observed that V_{L1} and V_{L2} differ from each other by a constant factor of 2 at every instant of time. Therefore, the two inductors can be coupled just as was discussed for the Cuk converter to make either the input current or the output current ripple zero.

Soft Switching in Converters

The switched-mode converter topologies consist of power semiconductor switches. They are in no way ideal switches as can be seen from the discussion on the various semiconductor switch characteristics in Chapter 1.

In general, most of the switches in the switched-mode power converters are not switched in excess of 100 kHz. Switching at higher frequencies in excess of 1 MHz is desirable as the isolating transformer can become smaller and the sizes of the output L and C can also become smaller, thereby improving or bringing down the volumetric power density (power per unit volume) of the converter. However, one of the most important factors which prevent switching at higher frequencies is the switching losses of the power devices. As discussed in Chapter 1, every switch dissipates power during switching transitions. Thus as the switching frequencies increases, there are more switching transitions in a given time and thereby more power loss. Therefore, it becomes important to tackle the issue of switching losses in the switched-mode converters.

The switching losses occur due to the increase of one variable (voltage or current) and the decrease of the other variable during the switching transition. The product of these two variables is non-zero during the switching transition which results in the switching loss. If one of the two variables, either the voltage across the switch or the current through the switch, is maintained at zero during the switching transition, then the product of the voltage and current is zero resulting in zero switching loss. Such zero switching loss converters are called *soft switching converters*. If the current is maintained at zero during the switching transition, then such soft switching converters come under the category of *ZCS converters*. On the other hand, if the voltage is maintained at zero during the switching transition, then such soft switching converters come under the category of *ZVS converters*.

Any of the converters discussed till now can be converted into ZCS or ZVS converter. The change is actually in the structure of the controlled switch. To realize ZCS, a primary requirement is that the controlled switch should have an inductor in series with it so that the current does not have discontinuities. Figure 5.60(a) shows the realization of a zero-current switch. The internal body diode of the controlled switch is needed for the operation of the switch.

To realize ZVS, a primary requirement is that the controlled switch should have a capacitor across it so that the voltage across the device does not have discontinuities. Figure 5.60(b) shows the realization of a zero-voltage switch. A diode D is included in series with the switch S. Most of the commercially available switches

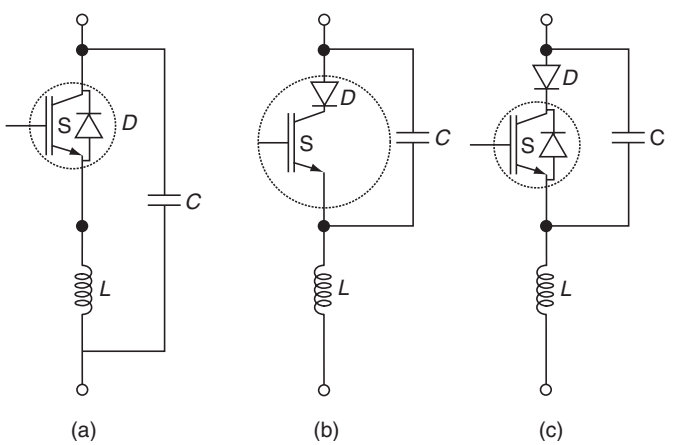


Figure 5.60 (a) Zero-current switch realization; (b) zero-voltage switch realization; (c) zero-voltage switch realization if internal body diode is present in the controlled switch.

like the MOSFETs and IGBTs will have internal body diodes. If these internal body diodes are present, they will automatically be blocked by the diode D as shown in Figure 5.60(c) to realize zero-voltage switch.

The controlled switches of any converter topology can be replaced with the zero-current switch or the zero-voltage switch as indicated in Figure 5.60 to obtain ZCS or ZVS operation of the converter. This principle can be demonstrated by replacing the controlled switch of the traditional buck converter with the zero-current switch and the zero-voltage switch to achieve ZCS and ZVS operations.

Zero-Current Switching Applied to Buck Converter

Figure 5.61 shows the circuit schematic of a buck converter with the controlled switch replaced with a zero-current switch. The switch S1 is an IGBT with an internal body diode D_1. L_s and C_s are components associated with the zero-current switch. L_s is the inductor that is in series with the switch S1 as per the zero-current switch requirements. C_s is a capacitor placed in parallel with the switch and inductor together as shown in Figure 5.61. The rest of the circuit is same as the buck converter circuit.

The following are the four operative modes of the ZCS converter:

1. load current freewheeling with controlled switch in OFF-state;
2. energy build up of ZCS inductor L_s at constant voltage;
3. L_s–C_s resonance operation;
4. capacitor C_s charging at constant current.

Figure 5.62 shows the four operative circuits in the four modes of operation of the ZCS converter. As the output inductor is very large compared to the soft-switching inductor L_s, I_o is considered to be constant during the switching period. The operation of the ZCS converter is described as follows after having reached the steady-state operating condition,

Mode 1 – S1 OFF and Load Current Freewheeling During this time, the output inductor current with average I_o will freewheel through the diode D_2 as shown. As D_2 is ON, one end of the capacitor C_s is at circuit ground potential and the other end is at V_i. Therefore, C_s will have a voltage of V_i across it with the polarity as shown in Figure 5.62(a).

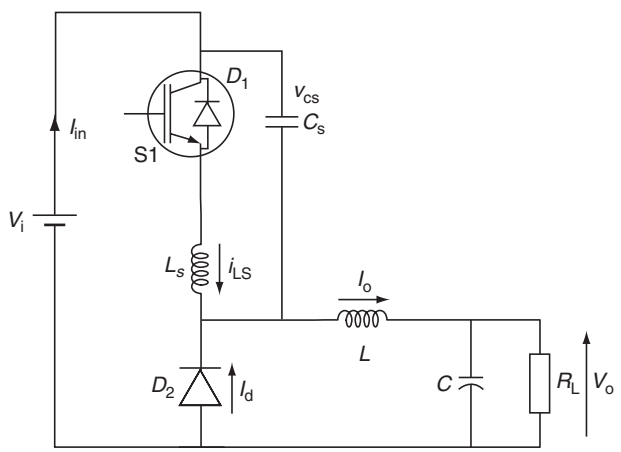


Figure 5.61 | Buck converter with zero-current switch.

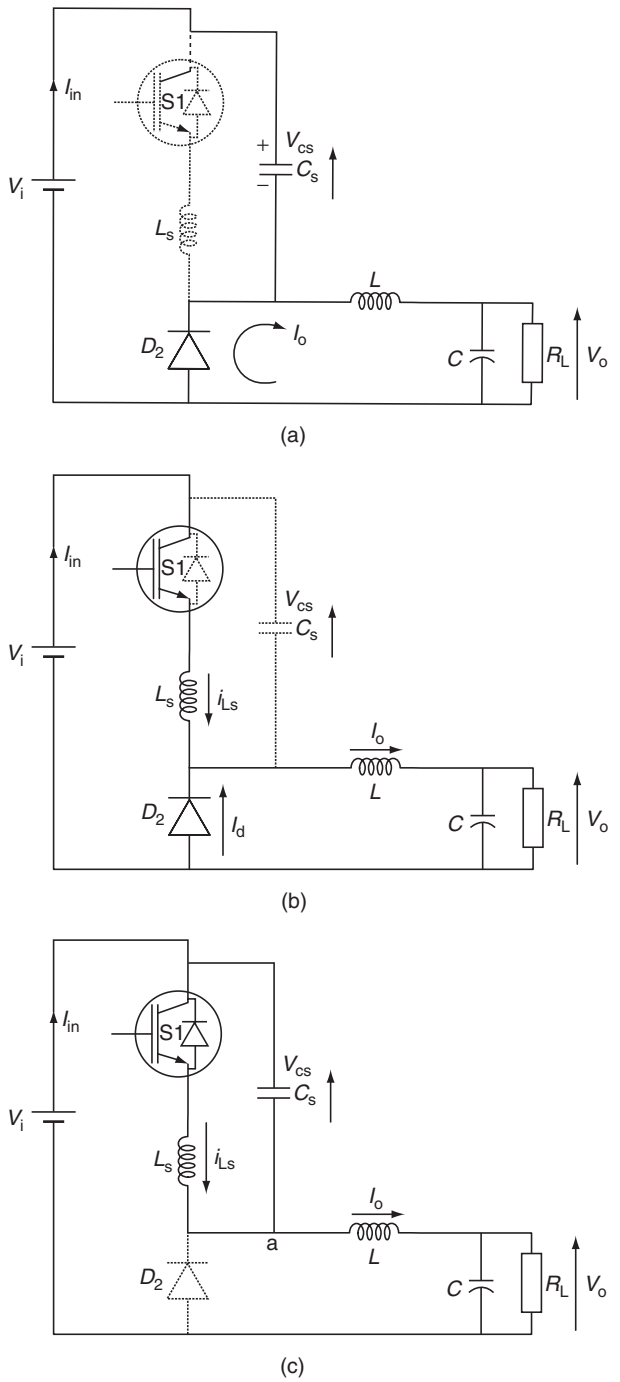
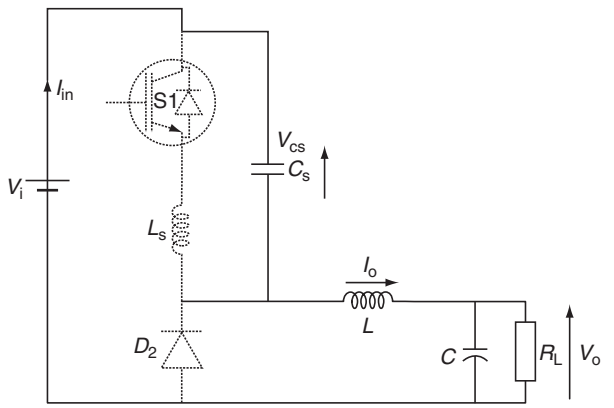


Figure 5.62 ZCS buck converter: (a) Load current freewheeling and S1 OFF; (b) energy build up of inductor L_s; (c) resonance operation; (d) capacitor C_s charging at constant current.



(d)

Figure 5.62 | *Continued*.

Mode 2 – S1 ON and Inductor Energy Build Up At time t_1, the switch S1 is turned ON. The operative circuit is shown in Figure 5.62(b). As D_2 is still conducting, the voltage across the inductor L_s is V_i. The inductor current i_{Ls} rises linearly with a rate of V_i/L_s. Thus,

$$i_{Ls} = \frac{V_i}{L_s}(t - t_1) \tag{5.112}$$

The current i_{Ls} rises linearly till it reaches the value of I_o when D_2 becomes reverse-biased and switches OFF.

Mode 3 – S1 ON and Resonance Operation The operative circuit is shown in Figure 5.62(c). After i_{Ls} reaches the value of I_o, D_2 is OFF. i_{Ls} will reach I_o at time t_2 that can be deduced from Eq. (5.112) and is given as

$$t_2 - t_1 = \frac{I_o L_s}{V_i} \tag{5.113}$$

C_s and L_s form a resonant tank. C_s discharges in a resonant manner through S1 and L_s. The current through the inductor L_s is the sum of the currents at the node "a" of Figure 5.62(c). Considering the time reference to start at the beginning of this mode, the inductor current is given as

$$i_{Ls} = I_o + \left(\frac{V_i}{Z}\right)\sin \omega t \tag{5.114}$$

where $Z = \sqrt{L_s / C_s}$ is the impedance of the resonant tank and $\omega = \sqrt{L_s C_s}$ is the natural frequency of resonance of the L_s–C_s tank. The voltage across the capacitor C_s is given as

$$v_{Cs} = V_i \cos \omega t \tag{5.115}$$

Observing the i_{Ls} waveform shown in Figure 5.63, it can be seen that the current waveform will resonate and cut the zero axis only if

$$\frac{V_i}{Z} > I_o \tag{5.116}$$

If t_3 is the time at which the inductor current i_{Ls} cuts the zero axis, then from Eq. (5.114), one obtains

$$0 = I_o + \left(\frac{V_i}{Z}\right)\sin\omega(t_3 - t_2)$$

$$t_3 - t_2 = \left(\frac{1}{\omega}\right)\left\{\pi + \sin^{-1}\left(\frac{I_o Z}{V_i}\right)\right\} \tag{5.117}$$

After the inductor current i_{Ls} reverses at t_3, the current flows through the internal body diode of the switch S1. During this time S1 can be turned OFF as the current through it is 0. Thus during both switch-ON and switch-OFF of S1, the current through it is 0 ensuring ZCS operation.

Mode 4 – S1 OFF and Capacitor Charging After the inductor current i_{Ls} reverses and flows through the body diode of the switch S1, the S1 is turned OFF. S1 is turned OFF anytime in the time band when i_{Ls} is negative and flowing through the body diode. The resonant current flow through the body diode of switch S1 continues till the inductor current i_{Ls} reaches zero again. Then the whole of the output current I_o will flow through the capacitor C_s. The capacitor C_s will now be charged to V_i at constant current I_o.

If t_4 is the moment when the current through the body diode and L_s becomes zero again, then from Eq. (5.114), one obtains

$$t_4 - t_2 = \left(\frac{1}{\omega}\right)\left\{2\pi - \sin^{-1}\left(\frac{I_o Z}{V_i}\right)\right\} \tag{5.118}$$

From t_4 onwards up to t_5 the capacitor charges linearly with constant current I_o. It reaches the voltage V_i at time t_5. The time t_5 can be calculated as

$$t_5 - t_4 = \frac{C_s(V_i - v_{Cs\text{-}t4})}{I_o} \tag{5.119}$$

where $v_{Cs\text{-}t4}$ is the voltage across C_s at time t_4 when the charging of the capacitor at constant current begins. It can be obtained from Eq. (5.115).

After the capacitor is charged to V_i, the current through the capacitor is zero and the I_o now flows through the freewheeling diode D_2. This is the same state as Mode 1. The whole cycle repeats.

Referring to the waveforms of Figure 5.63, it can be observed that the output voltage is the average of the V_{D2} waveform. Thus the input–output relationship is not a simple linear function of the duty ratio as in the case of the traditional buck converter, but it is a non-linear function of the switching frequency and the resonant frequency. It should also be noted that the time duration of the ON-state of the switch S1 is

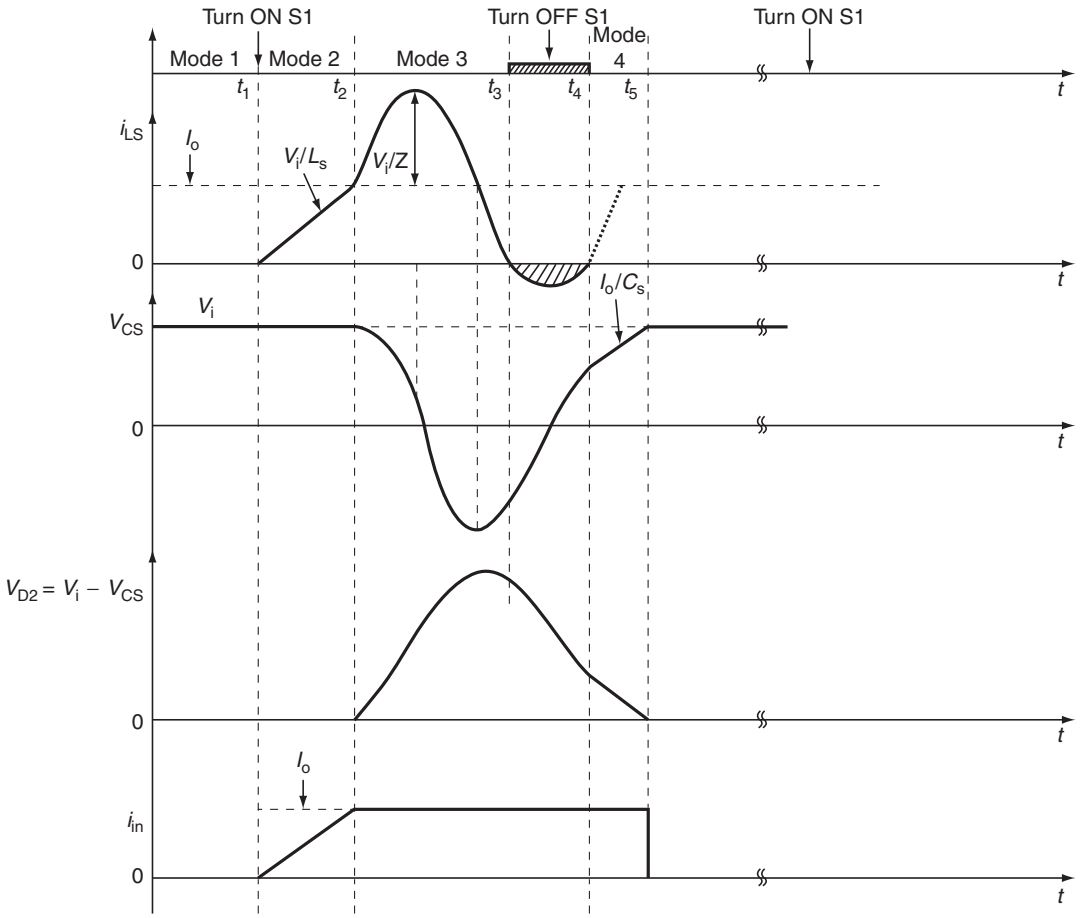


Figure 5.63 | Waveforms of important signals of ZCS buck converter.

decided by the resonant circuit of the zero-current switch. But the time duration of the OFF-state of the switch S1 can be controlled to any length.

One of the main drawbacks of the ZCS topology is that at very high frequency the parasitic capacitance across the switch S1 becomes significant and its discharge current through S1 cannot be limited on any account. Therefore, ZCS topology has an upper frequency bound beyond which it does not serve the purpose of lossless switching.

Zero-Voltage Switching Applied to Buck Converter

Figure 5.64 shows the circuit schematic of a buck converter with the controller switch replaced with a zero-voltage switch. The switch S1 is an IGBT. L_s and C_s are components associated with the zero-voltage switch. C_s is the capacitor that is in parallel across the switch S1 as per the zero-voltage switch requirements. L_s is an inductor placed in series with the switch and capacitor together as shown in Figure 5.64. The rest of the circuit is same as the buck converter circuit.

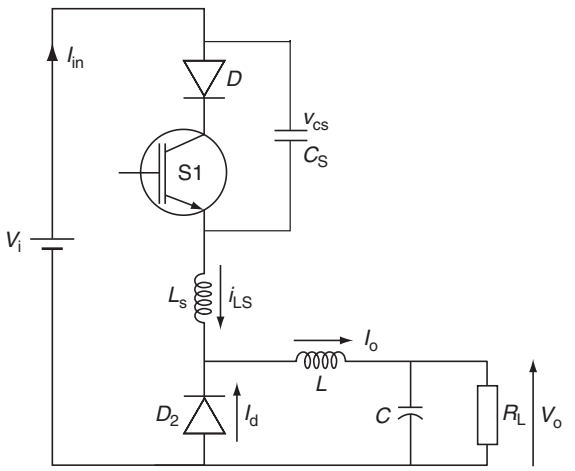


Figure 5.64 | Schematic of the ZVS buck converter.

The following are four operative modes of the ZVS converter:

1. load current through switch S1;
2. capacitor C_s charging at constant current;
3. L_s–C_s resonance operation;
4. energy build up of ZVS inductor L_s at constant voltage.

Figure 5.65 shows the four operative circuits in the four modes of operation of the ZVS converter. As the output inductor is very large compared to the soft-switching inductor L_s, I_o is considered to be constant during the switching period. The operation of the ZVS converter is described as follows after having reached the steady-state operating condition.

Mode 1 – S1 ON and Load Current through S1 During this time, the output inductor current with average I_o will flow through S1 as shown. As S1 is ON, the voltage across C_s is zero. The current through L_s is same as the current flowing through S1, that is, I_o.

Mode 2 – S1 OFF and Capacitor Charging At time t_1, the switch S1 is turned OFF. Due to the presence of C_s the voltage across the switch S1 is zero and will not change suddenly. The operative circuit is shown in Figure 5.65(b). Once the switch S1 turns OFF, the inductor current now shifts to C_s which starts charging at constant current I_o. The capacitor voltage v_{Cs} rises linearly with a rate of I_o/C_s. Thus,

$$v_{Cs} = \frac{I_o}{C_s}(t - t_1) \tag{5.120}$$

The voltage v_{Cs} rises linearly till it reaches the value of V_i when D_2 becomes forward-biased and switches ON.

Mode 3 – S1 OFF and Resonance Operation The operative circuit is shown in Figure 5.65(c). After v_{Cs} reaches the value of V_i, D_2 is ON. The value of v_{Cs} will reach that of V_i at time t_2 that can be deduced from Eq. (5.120) and is given as

$$t_2 - t_1 = \frac{V_i C_s}{I_o} \tag{5.121}$$

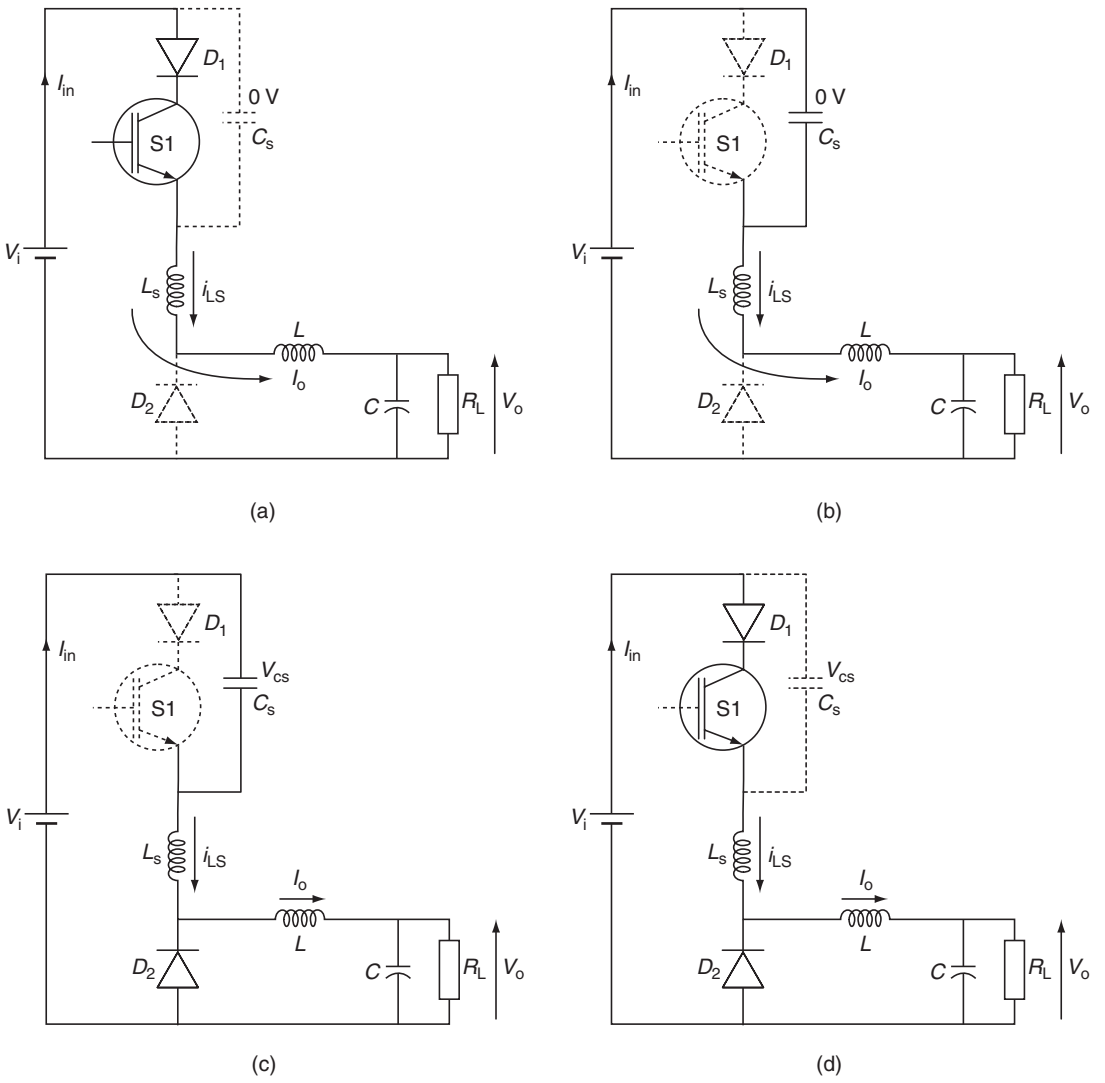


Figure 5.65 ZVS buck converter: (a) Load current through S1; (b) capacitor C_s charging at constant current; (c) resonance operation; (d) energy build up of inductor L_s.

C_s and L_s form a resonant circuit. L_s current falls in a resonant manner through D_2 and C_s. Applying Kirchhoff's voltage law in the resonant loop shown in Figure 5.65(c) and considering the time reference to start at the beginning of this mode, the capacitor voltage is given as

$$v_{Cs} = V_i + (I_o Z)\sin \omega t \tag{5.122}$$

where $Z = \sqrt{L_s / C_s}$ is the impedance of the resonant tank and $\omega = \sqrt{L_s C_s}$ is the natural frequency of resonance of the L_s–C_s tank.

The current through the inductor L_s is given as

$$i_{Ls} = I_o \cos \omega t \tag{5.123}$$

Observing the v_{Cs} waveform shown in Figure 5.66, it can be seen that the voltage waveform will resonate and cut the zero axis only if

$$I_o Z > V_i \tag{5.124}$$

If t_3 is the time at which the capacitor voltage v_{Cs} cuts the zero axis, then from Eq. (5.122), one obtains

$$0 = V_i + (I_o Z) \sin \omega (t_3 - t_2)$$

$$t_3 - t_2 = \left(\frac{1}{\omega}\right)\left\{\pi + \sin^{-1}\left(\frac{V_i}{I_o Z}\right)\right\} \tag{5.125}$$

After the capacitor voltage v_{Cs} reverses at t_3, the switch S1 can be turned ON again.

Mode 4 – S1 ON and Inductor Energy Build Up After the capacitor voltage v_{Cs} reverses the switch S1 is turned ON. S1 is turned ON anytime in the time band when v_{Cs} is negative. The negative voltage will be supported by the diode D. The resonant current flows through D_2, L_s and C_s till the capacitor voltage v_{Cs}

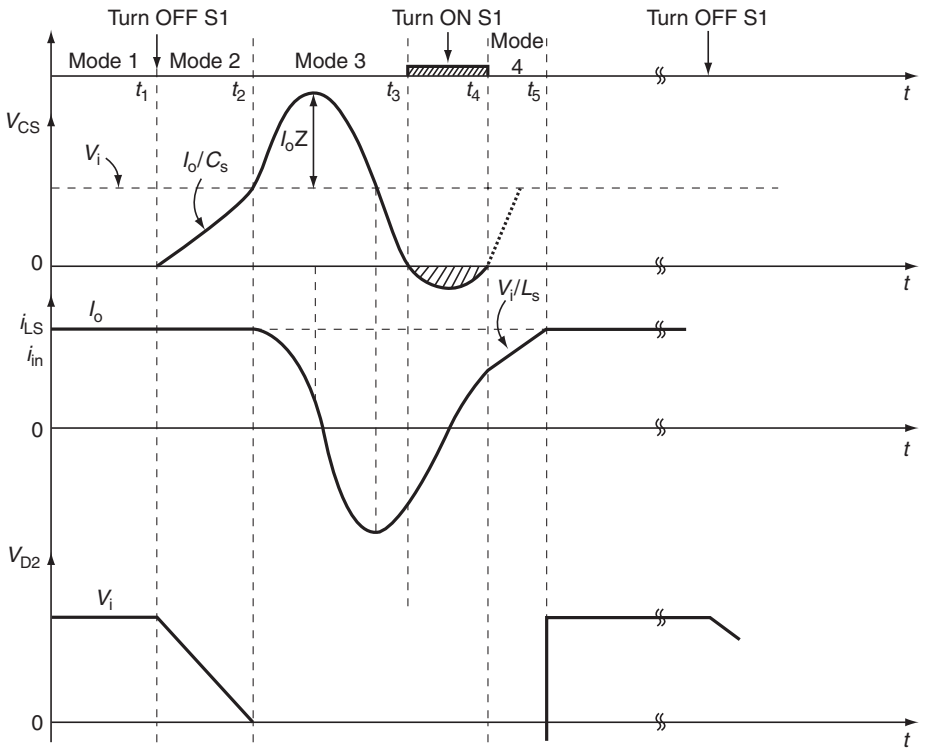


Figure 5.66 | Waveforms of important signals of ZVS buck converter.

reaches zero. Then the whole input voltage V_i will appear across L_s. The energy in L_s will now build up to I_o at constant voltage V_i across it.

If t_4 is the moment when the voltage across the capacitor reaches zero again, then from Eq. (5.122), one obtains

$$t_4 - t_2 = \left(\frac{1}{\omega}\right)\left\{2\pi - \sin^{-1}\left(\frac{V_i}{I_o Z}\right)\right\} \tag{5.126}$$

From t_4 onwards up to t_5 the inductor energy builds up linearly with constant voltage V_i across it. It reaches the current I_o at time t_5. The time t_5 can be calculated as

$$t_5 - t_4 = \frac{L_s(I_o - i_{Ls\text{-}t4})}{V_i} \tag{5.127}$$

where $i_{Ls\text{-}t4}$ is the inductor current through L_s at time t_4 when the inductor energy build up at constant voltage begins. The value $i_{Ls\text{-}t4}$ can be found from Eq. (5.123).

After the inductor current has build up to I_o, the converter is in Mode 1. The whole cycle repeats.

Referring to the waveforms of Figure 5.66, it can be observed that the output voltage is the average of the V_{D2} waveform. It should also be noted that the time duration of the OFF-state of the switch S1 is decided by the resonant circuit of the zero-voltage switch. However, the time duration of the ON-state of the switch S1 can be controlled to any length. It should also be noted that as the device parasitic capacitance can be taken as part of the resonating capacitor C_s; the ZVS topology can be operated at much higher frequencies as compared to the ZCS topology.

The ZCS and ZVS converters are also called quasi-resonant converters as L_s and C_s resonate during a fraction of the switching period.

CONCLUDING REMARKS

The DC–DC converters are one of the most widely used circuits in electronic equipments. Any electronic equipment requires one or more power supplies for its operation. Therefore, whether it is low-power applications like the mobile phones, digital camera, battery chargers or high-power applications like motor drive controller, uninterruptible power system, etc., DC–DC converters are used to act as power interface between various DC buses.

The converters discussed in this chapter are the more popular ones. However, there are numerous topologies in the literature for specific applications. The key point to be stressed here is that the steady-state analysis gives a framework for designing the open-loop converter system. The static *i–v* characteristics of the devices and the SPDT switch requirements should be matched to arrive at the best power semiconductor device that may be used as a switch for a given converter topology. Among the isolated power converters, the flyback converter is by far the most popular. This is because the flyback converter has the least component count and simple to implement.

With MOSFETs as the power switch, switching frequencies of 100 kHz are easily achievable. However, to reduce the size of the magnetics (inductor and transformer cores) one may need to switch to higher frequencies that tend towards 1 MHz. Also, the MOSFET switching losses and the core losses could become prohibitively high. Towards this end, soft switching may be introduced to decrease the switching losses.

The power supplies are in general multiple-output winding converters that supply multiple isolated outputs. The regulation aspects and the

controller issues with respect to control of multiple outputs are addressed in Chapter 10. Finally, the laboratory exercises given in the next section are meant to improve the insights gained in this chapter and form the first steps towards practical implementation of full-fledged DC–DC power converters.

LABORATORY EXERCISES

1. A pulse-width modulated (PWM) waveform is required to operate the switches of the DC–DC converter. There are numerous ways in which a PWM waveform may be generated. It can be generated either by discretely building circuits for generating a triangular carrier and comparing a DC waveform with the carrier or by using a digital port of the microcontroller. There are also many commercial ICs available that are dedicated to generate PWM waveforms. A sample PWM generation circuit using TL494 IC is shown in Figure 5.67.

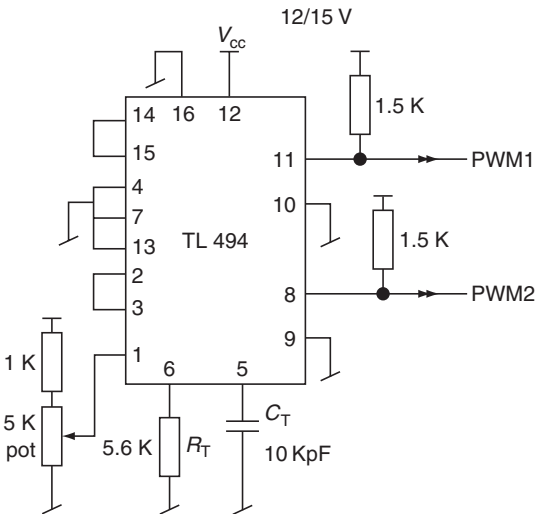


Figure 5.67 | Pulse-width modulator circuit.

Mode of implementation: The above circuit is studied by

a. Hardware bread-boarding

Tasks for study:

(a) Rig up the circuit as shown in Figure 5.67.

(b) View the waveform at pin 5. What is the nature of the waveform? Is it a sawtooth waveform? Measure the extremities on the *y*-axis and also measure the frequency of the waveform. Check from the nomograph in the datasheet of the IC. Observe whether or not the frequency obtained is correct.

(c) Change R_T and C_T and observe the change in the frequency.

(d) Measure the voltage at pin 1. Adjust the potentiometer till the voltage at pin 1 is at a value in between the *y*-axis extremities of sawtooth voltage at pin 5.

(e) Monitor the voltage at pins 11 and 8.

(f) What happens when the voltage at pin 1 is changed? Why?

(g) What is the voltage at pin 14?

(h) What is the function of pin 13? Check with the datasheet and discuss its application.

(i) Why are pins 15 and 16 connected in the manner shown in Figure 5.67? What other ways can pins 15 and 16 be used?

(j) Resistors of value 1.5 kΩ are connected from pins 8 and 11 to V_{cc}. What is the lowest value to which these resistors may be reduced? Can they directly drive BJTs and MOSFETs? Discuss.

2. The circuit shown in Figure 5.68 is a buck converter circuit or a step-down DC–DC converter. The transistor or BJT is operated as a switch, that is, it is either in ON-state or OFF-state. Instead of a BJT, one may use a MOSFET or IGBT.

Mode of implementation: The above circuit is studied by

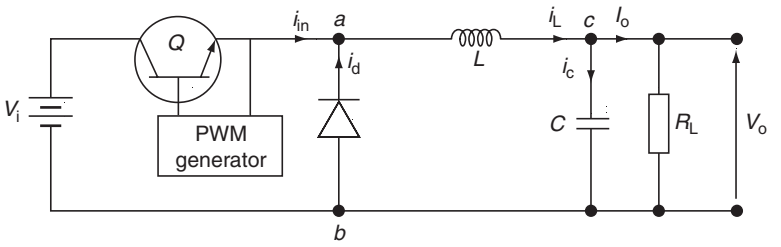


Figure 5.68 | Buck converter circuit.

a. Simulation in SciLab
b. Simulation in SPICE
c. Hardware bread-boarding

Tasks for study:

(a) Rig up the above circuit with appropriate values for the components.
(b) Observe the voltage waveforms across the diode, inductor and output.
(c) What is the average voltage waveform across the inductor during steady state?
(d) Observe the inductor current and the output voltage waveforms together. Why does the output voltage ripple waveshape differ significantly from the triangular inductor current ripple waveshape?
(e) Tabulate the steady-state input and output voltages for various duty ratios D. From the tabulation, find the relationship between duty ratio D, input voltage and output voltage.
(f) Tabulate the steady-state input and output currents for various duty ratios D. From the tabulation, find the relationship between duty ratio D, input current and output current?
(g) What is the effect of load variation on the inductor current ripple amplitude?
(h) What happens to the inductor current as the load is decreased? What is the minimum load to maintain continuous current in the inductor?
(i) Tabulate and plot V_o/V_i versus the duty ratio for the converter. Plot duty ratio on the x-axis and V_o/V_i on the y-axis. Based on the plot, specify the range of valid duty ratio for the converter.
(j) What is the value of the conduction parameter? Plot K_{bound} versus d, $1-d$, $(1-d)^2$ and $d(1-d)^2$ for the converter. From the plots, comment on the value of K_{bound} for the converter.

3. The circuit shown in Figure 5.69 is a buck converter circuit. The specifications for this circuit are: $V_{in} = 15$ V; $V_o = 5$ V; $I_o = 50$ mA – 500 mA; $f_s = 20$ kHz; $\Delta V_o = 50$ mV.

Mode of implementation: The above circuit is studied by

a. Hardware bread-boarding

Tasks for study:

(a) Rig up the above circuit with appropriate values for the components.
(b) The inductor L has to be fabricated by winding copper wire on an EE-10 ferrite core. A paper may be inserted in the central limb of the EE core as air gap to obtain the proper inductor value of the inductor. To experimentally measure the inductor value, a small test jig has to be built using a MOSFET switch with the test inductor connected between the positive rail and the drain of the MOSFET. A freewheeling diode–resistor should be connected across the inductor. The current through the inductor can be measured by including a very small resistor in the source of the

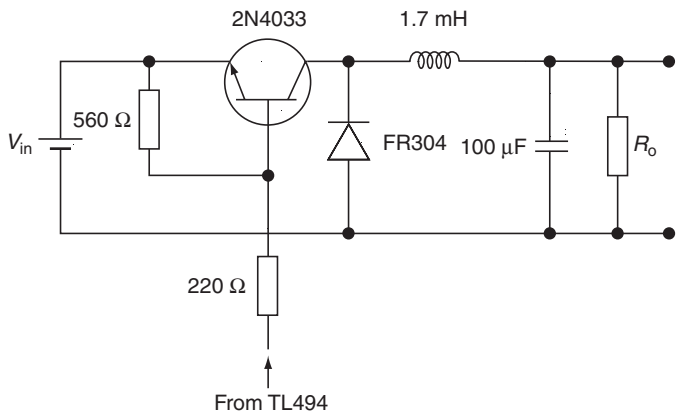


Figure 5.69 | Buck converter circuit.

switch. Switching pulses are given to the gate of the MOSFET and the voltage across the inductor and the current through the inductor are measured in an oscilloscope. Using Faraday's equation, calculate the value of the inductor.

(c) Observe the voltage waveforms across the diode, inductor and output.

(d) Observe the currents through L, switch and the load.

(e) Tabulate the steady-state input and output voltages for various duty ratios D. From the tabulation, find the input–output voltage relationship and compare with the idealized input–output voltage relationship.

(f) What is the effect of load variation on the inductor current ripple amplitude?

(g) What happens to the inductor current as the load is decreased? What is the minimum load to maintain continuous current in the inductor?

(h) Tabulate and plot V_o/V_{in} versus the duty ratio for the converter. Plot duty ratio on the x-axis and V_o/V_{in} on the y-axis. Based on the plot, specify the range of valid duty ratio for the converter.

(i) Find out the value of K for the circuit. Plot K_{bound} versus d for the converter.

4. The circuit shown in Figure 5.70 is a boost converter circuit or a step-up DC–DC converter. The transistor or BJT is operated as a switch, that is, it is either in ON-state or OFF-state. Instead of a BJT, one may use a MOSFET or IGBT.

Mode of implementation: The above circuit is studied by

a. Simulation in SciLab

b. Simulation in SPICE

c. Hardware bread-boarding

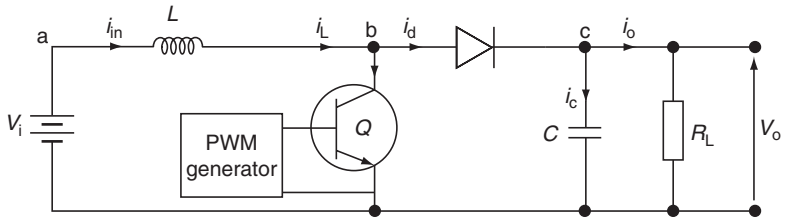


Figure 5.70 | Boost converter circuit.

Tasks for study:

(a) Rig up the above circuit with appropriate values for the components.
(b) Observe the voltage waveforms across the diode, inductor and output.
(c) What is the average voltage waveform across the inductor during steady state?
(d) Observe the inductor current and the output voltage waveforms together.
(e) Tabulate the steady-state input and output voltages for various duty ratios D. From the tabulation, find the relationship between duty ratio D, input voltage and output voltage.
(f) Tabulate the steady-state input and output currents for various duty ratios D. From the tabulation, find the relationship between duty ratio D, input current and the output current.
(g) What is the effect of load variation on the inductor current ripple amplitude?
(h) What happens to the inductor current as the load is decreased? What is the minimum load to maintain continuous current in the inductor?
(i) Tabulate and plot V_o/V_i versus the duty ratio for the converter. Plot duty ratio on the x-axis and V_o/V_i on the y-axis. Based on the plot, specify the range of valid duty ratio for the converter.
(j) What is the value of the conduction parameter? Plot K_{bound} versus d, $1 - d$, $(1 - d)^2$ and $d(1 - d)^2$ for the converter. From the plots, comment on the value of K_{bound} for the converter.

5. The circuit shown in Figure 5.71 is a buck–boost converter circuit. The transistor or BJT is operated as a switch, that is, it is either in ON-state or OFF-state. Instead of a BJT, one may use a MOSFET or IGBT.

Mode of implementation: The above circuit is studied by

a. Simulation in SciLab
b. Simulation in SPICE
c. Hardware bread-boarding

Tasks for study:

(a) Rig up the above circuit with appropriate values for the components.
(b) Observe the voltage waveforms across the diode, inductor and output.
(c) What is the average voltage waveform across the inductor during steady state?
(d) Observe the inductor current and the output voltage waveforms together.
(e) Tabulate the steady-state input and output voltages for various duty ratios D. From the tabulation, find the relationship between duty ratio D, input voltage and output voltage.
(f) Tabulate the steady-state input and output currents for various duty ratios D. From the tabulation, find the relationship between duty ratio D, input current and output current.

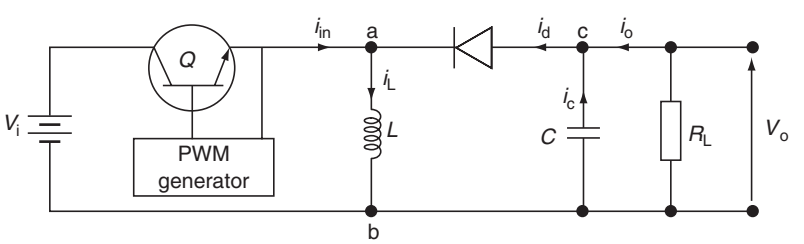


Figure 5.71 | Buck–boost converter circuit.

(g) What is the effect of load variation on the inductor current ripple amplitude?

(h) What happens to the inductor current as the load is decreased? What is the minimum load to maintain continuous current in the inductor?

(i) Tabulate and plot V_o/V_i versus the duty ratio for the converter. Plot duty ratio on the x-axis and V_o/V_i on the y-axis. Based on the plot, specify the range of valid duty ratio for the converter.

(j) What is the value of the conduction parameter? Plot K_{bound} versus d, $1 - d$, $(1 - d)^2$ and $d(1 - d)^2$ for the converter. From the plots, comment on the value of K_{bound} for the converter.

6. The circuit shown in Figure 5.72 is a forward converter circuit. The transistor or BJT is operated as a switch, that is, it is either in ON-state or OFF-state. Instead of a BJT, one may use a MOSFET or IGBT.

Mode of implementation: The above circuit is studied by

a. Simulation in SciLab

b. Simulation in SPICE

c. Hardware bread-boarding

Tasks for study:

(a) Rig up the above circuit with appropriate values for the components.

(b) Observe the voltage waveforms across the BJT, primary, secondary, the secondary freewheeling diode and output.

(c) Observe the current waveforms through the various components of the circuit.

(d) Keeping the duty cycle at 0.4, monitor the voltage waveform across the BJT. What is the effect of duty cycle on the waveshape? What should be the voltage rating of the BJT?

(e) For a given load, I_o, what should be the current rating of the BJT?

(f) Why should duty cycle be less than 0.5? What happens if duty cycle is greater than 0.5?

(g) What is the primary current waveshape for duty cycle of 0.75? What will be the primary current waveshape for duty cycle of 0.75 if a practical transformer is used?

(h) For a given load, I_o, what should be the current ratings of the various diodes?

(i) What is the input–output voltage and current transfer ratios under steady state?

7. The circuit shown in Figure 5.73 is a flyback converter circuit. The transistor or BJT is

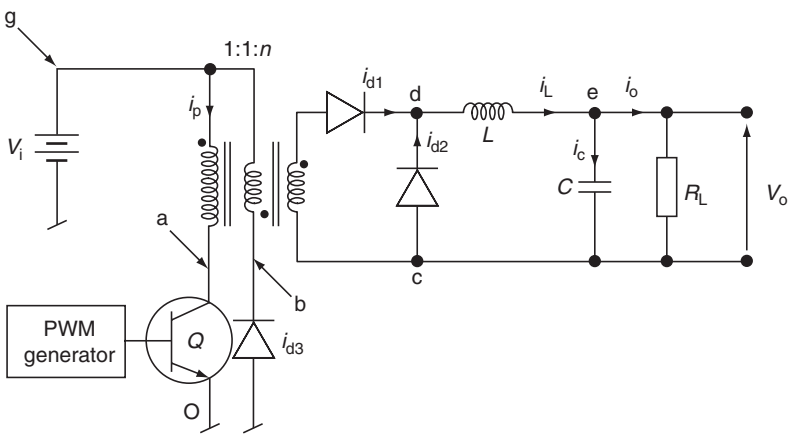


Figure 5.72 | Forward converter circuit.

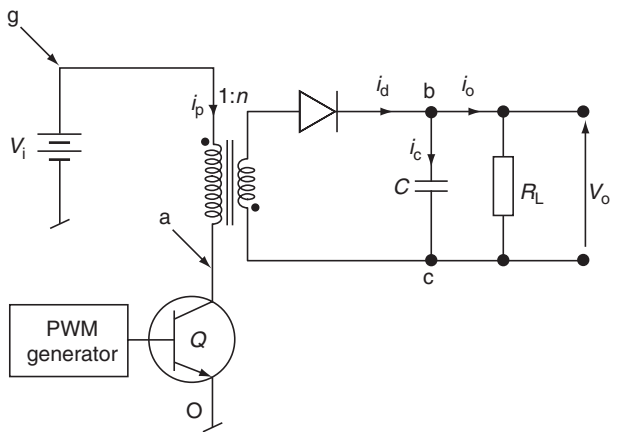


Figure 5.73 | Flyback converter circuit.

operated as a switch, that is, it is either in ON-state or OFF-state. Instead of a BJT, one may use a MOSFET or IGBT.

Mode of implementation: The above circuit is studied by

- **a.** Simulation in SciLab
- **b.** Simulation in SPICE
- **c.** Hardware bread-boarding

Tasks for study:

- **(a)** Rig up the above circuit with appropriate values for the components.
- **(b)** Observe the voltage waveforms across the BJT, primary, secondary and output.
- **(c)** Observe the current waveforms through the various components of the circuit.
- **(d)** Keeping the duty cycle at 0.4, monitor the voltage waveform across the BJT. What is the effect of duty cycle on the waveshape?
- **(e)** For a given load, I_o, what should be the current rating of the BJT?
- **(f)** Is there a duty cycle limitation of 0.5? What happens if duty cycle is greater than 0.5?
- **(g)** Does increase in load cause increase in the primary current?
- **(h)** As current flow in primary and secondary is mutually exclusive in time, how does loading affect the primary current and hence primary power?
- **(i)** What is the transformer turns ratio?
- **(j)** What is the input–output voltage and current transfer ratios under steady state?

FILL IN THE BLANKS

1. A chopper can be constructed using an _____ switch.
2. The chopper does not contain an output _____.
3. A single basic chopper consisting of the single pole double throw switch can handle only _____ quadrants of operation.
4. To handle all four quadrant operation,_____ basic choppers are required.
5. In any converter topology, the inductor should be placed in such a manner where its current will not become _____ at instant of operation.
6. In any converter topology, the capacitor should be placed in such a manner that its _____

does not become discontinuous at instant of operation.

7. The buck converter gives an output voltage that is always _____ than the input voltage.

8. In a buck converter, the inductor current is composed of the DC _____ current part and the AC _____ current part.

9. The capacitor value is calculated by applying the _____ rule.

10. The area under the positive portion of the capacitor current waveform implies _____ of the capacitor.

11. The area under the negative portion of the capacitor current waveform implies _____ of the capacitor.

12. The boost converter gives an output voltage that is _____ than the input voltage.

13. The inductor current is composed of _____ which is the DC _____current part and _____ which is the AC _____ part.

14. The diode current is composed of the DC _____ current part and the AC _____ current part.

15. The buck–boost converter gives an output voltage that is either _____ or _____ than the input voltage.

16. The inductor current is composed of the _____ values of the input and output current as the DC part and Δi_L as the _____ part.

17. The diode current is composed of the DC _____ current part and the AC _____ current part.

18. The switch element in any converter is selected based on the _____ *i–v* static characteristic requirement.

19. Continuous conduction mode and discontinuous conduction mode are defined with respect to the _____.

20. If the current through the inductor does not become zero at any instant in a switching period, then the converter operation is referred to as the _____.

21. If the current through the inductor becomes zero for some portion of the switching period then the converter operation is referred to as the _____.

22. If the input and output voltages at any given instant are constant, then the inductor ripple current amplitude is _____ with load.

23. For a particular converter, if the conduction parameter value is greater than K_{bound} for the converter at a given operating duty cycle, then the converter is in _____.

24. For a particular converter, if the conduction parameter value is lesser than K_{bound} for the converter at a given operating duty cycle, then the converter is in _____.

25. The buck-based isolated topologies consist of the output _____ converter, the input side _____ and the _____ transformer.

26. In isolated converters, the _____ circuit is needed to convert the DC input source to an AC pulse voltage before feeding to the transformer.

27. In the forward converter, the flux in the core increases _____ during the DT_s duration of the switching period.

28. In the forward converter with demagnetizing winding the leakage inductance is minimized by winding the primary and the demagnetizing windings in a _____ fashion.

29. In the forward converter with primary and demagnetizing windings having equal number of turns, the maximum duty cycle is _____.

30. The forward converter utilizes only the _____ half of the core magnetization as the magnetizing current and the core flux are _____.

31. The push–pull, half-bridge and full-bridge converters magnetize the core in both _____ directions to better utilize the core.

32. The switching frequency of the primary side switches is _____ the frequency of the inductor ripple current.

33. In the half-bridge converter, the voltage across the primary of the transformer is _____ the DC link voltage.

34. The voltage developed across the OFF switches of the half-bridge converter is maximum _____ voltage.

35. In the full-bridge converter, the voltage across the primary of the transformer is the _____.

36. The voltage developed across the OFF switches of the full-bridge converter is maximum _____ voltage.

37. For a given power, the average current flowing through the primary of the full-bridge converter is _____ of that flowing through the half-bridge converter.

38. The current-carrying capacity for the full-bridge converter switches is _____ that of the switches of the half-bridge converter of the same output power rating.

39. Flyback converter is an isolated converter that is based on the _____ converter topology.

40. The flyback converter has one of the _____ component count.

41. In the high-boost converter, at a duty cycle of 0.5, the output voltage is _____.

42. In soft switching converters, the switching loss is _____.

43. The ZCS and ZVS converters are called also called _____ resonant converters.

DESCRIPTIVE QUESTIONS

1. What is duty ratio or duty cycle?

2. Can a chopper handle bi-directional power flow? Illustrate with an application example.

3. Explain the operation of the dual chopper.

4. What is volt-second balance?

5. What is charge balance?

6. Is volt-second balance and charge balance applicable under transient conditions? Explain.

7. Explain the operation of the buck converter and illustrate the operation with the inductor current and the switch waveforms.

8. What is inductor current freewheeling? Why does it occur?

9. Figure 5.13 gives the buck converter realizations. Can they be used for two-quadrant operation wherein the power flow can be bi-directional? Explain.

10. In a buck converter, what is the rate at which the inductor current ripple rises and falls?

11. Consider the buck converter circuit schematic of Figure 5.14 wherein the first-level non-idealities are included. What is the input–output relation? By what factor is the input–output relation different from the input–output relationship of the ideal buck converter? What is the efficiency of the buck converter circuit with non-idealities?

12. Explain the operation of the boost converter and illustrate the operation with the inductor current and the switch waveforms.

13. Figure 5.17 gives the boost converter realizations. Can they be used for two-quadrant

operation wherein the power flow can be bi-directional? Explain.

14. In a boost converter, what is the rate at which the inductor current ripple rises and falls?

15. Consider the boost converter circuit schematic of Figure 5.18 wherein the first-level non-idealities are included. What is the input–output relation? By what factor is the input–output relation different from the input–output relationship of the ideal boost converter? What is the efficiency of the boost converter circuit with non-idealities?

16. Explain the operation of the buck–boost converter and illustrate the operation with the inductor current and the switch waveforms.

17. Figure 5.21 gives the buck–boost converter realizations. Can they be used for two-quadrant operation wherein the power flow can be bi-directional? Explain.

18. Consider the buck–boost converter circuit schematic of Figure 5.22 wherein the first-level non-idealities are included. What is the input–output relation? By what factor is the input–output relation different from the input–output relationship of the ideal buck–boost converter? What is the efficiency of the buck–boost converter circuit with non-idealities?

19. For the DC–DC converter topology shown in Figure 5.74, switch S1 is connected to the input during DT_s duration and switch S2 is connected to the output during $(1 - D)T_s$ duration of a switching period. N_1 and N_2 are the number of turns on the tapped inductors L_1 and L_2 wound on the same core. Using the volt-second and charge balance rule, derive the steady-state input–output relationship. Plot the idealized *i–v* characteristic requirement for the switches S1 and S2 and indicate the most appropriate power semiconductor switches based on the static characteristic.

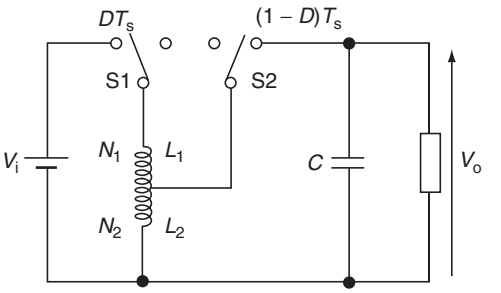


Figure 5.74 DC–DC converter topology for Question 19.

20. For the DC–DC converter topology shown in Figure 5.75, switch S1 is connected to the input during DT_s duration and switch S2 is connected to the output during $(1 - D)T_s$ duration of a switching period. N_1 and N_2 are the number of turns on the tapped inductors L_1 and L_2 wound on the same core. Using the volt-second and charge balance rule, derive the steady-state input–output relationship. Plot the idealized *i–v* characteristic requirement for the switches S1 and S2 and indicate the most appropriate power semiconductor switches based on the static characteristic.

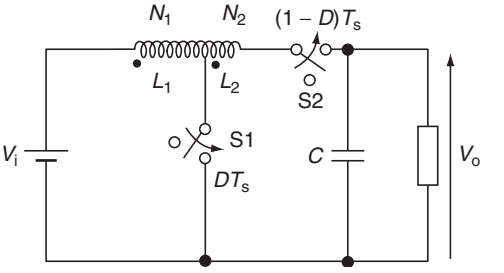


Figure 5.75 DC–DC converter topology for Question 20.

21. For the DC–DC converter topology shown in Figure 5.76, a single pole triple throw switch (SPTT) is connected as shown. In a switching period, the pole P is connect to throw T1 for a duration D_1T_s and is connected to throw T2

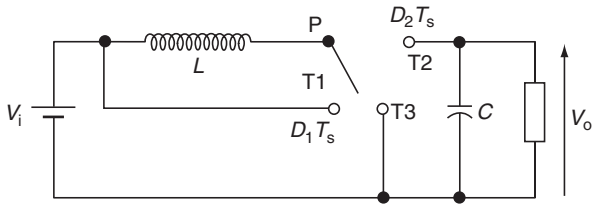


Figure 5.76 | DC–DC converter topology of Question 21.

for a duration D_2T_s. For the remaining duration of the switching period the pole P is connected to throw T3. Using the volt-second and charge balance rule, derive the steady-state input–output relationship. Plot the idealized i–v characteristic requirement for the switches of the SPTT switch and indicate the most appropriate power semiconductor switches based on the static characteristic.

22. What is conduction parameter? What is its significance?

23. Derive K_{bound} for the buck, boost and buck–boost converters and plot K_{bound} versus the duty cycle.

24. Derive the input–output relationship for a buck converter operating in DCM.

25. Derive the input–output relationship for a boost converter operating in DCM.

26. Derive the input–output relationship for a buck–boost converter operating in DCM.

27. Why is isolation needed between the output circuit and the input circuit?

28. What are the categories of isolated converters?

29. What are the types of buck-based isolated converters?

30. In forward converter, what is meant by core flux resetting?

31. What are the various methods of achieving core flux resetting?

32. Explain the operation of the forward converter and illustrate with inductor current, transformer primary, switch current and switch voltage waveforms.

33. For a forward converter, include the first-level non-idealities like the ON-state drops of the switches, transformer leakage, winding resistances and equivalent series resistance of the output capacitor. Derive the input–output relationship using the volt-second balance and charge balance rule.

34. What are the advantages of the forward converter with demagnetizing winding over the diode–resistor reset topology?

35. Explain the operation of the forward converter with demagnetizing winding.

36. What are the advantages of the dual switch forward converter and explain its functioning?

37. Explain the output inductor freewheeling process in the case of the push–pull, half-bridge and full-bridge converters. Draw the secondary side diode current waveforms.

38. What is flux walking in push–pull circuit? What is its effect in the functioning of the circuit?

39. Explain the operation of the half-bridge topology.

40. Explain the operation of the full-bridge topology.

41. How is flux walking problem solved in half-bridge and full-bridge converters?
42. Discuss the operation of the Cuk converter.
43. What are the benefits and drawbacks of the Cuk converter with respect to the buck-based converters?
44. How does one achieve zero ripple at the input or output inductor currents of the Cuk converter?
45. Can both the input and the output inductor current ripples be made zero simultaneously? Explain.
46. What conditions should be satisfied for coupling the inductors of a converter onto to a common core?
47. Discuss the operation of the high-boost converter.
48. Discuss the voltage rating of the IGBT and diode switches of a high-boost converter.
49. Can the inductors of the high-boost converter be coupled together onto a common core? Discuss.
50. What is the motivation for using soft switching in converters?
51. What are zero-current switching and zero-voltage switching?
52. Explain the ZCS and ZVS switch topologies.
53. Explain the operation of ZCS and ZVS for boost and buck–boost converters.
54. What are the four operative modes of the ZCS buck converter?
55. What are the four operative modes of the ZVS buck converter?

PROBLEMS

1. The input to a chopper is from a 100 V DC source. The chopper is switched at a frequency of 100 kHz with a pulse width of 4 μs. What is the average output voltage of the chopper?
2. A buck converter is supplied with an input voltage that varies between 20 V and 30 V. The output is required to be regulated at 5 V. Find the duty cycle range.
3. A buck converter that is switching at 50 kHz is supplied with an input voltage that varies between 20 V and 30 V. The output is required to be regulated at 5 V. A load resistor of 5 Ω is connected across the output. If the maximum allowable inductor current ripple is 10% of the load current, estimate the value of the inductance to be used in the buck converter.
4. For Problem 3, if the output voltage ripple specification is 1% of the output voltage, estimate the output capacitor value.
5. A boost converter is supplied with an input voltage that varies between 5 V and 10 V. The output is required to be regulated at 15 V. Find the duty cycle range.
6. A boost converter that is switching at 50 kHz is supplied with an input voltage that varies between 5 V and 10 V. The output is required to be regulated at 15 V. A load resistor of 15 Ω is connected across the output. If the maximum allowable inductor current ripple is 10% of the average inductor current, estimate the value of the inductance to be used in the boost converter.
7. For Problem 6, if the output voltage ripple specification is 1% of the output voltage, estimate the output capacitor value.
8. A buck–boost converter is supplied with an input voltage that varies between 5 V and 10 V. The output is required to be regulated at 15 V. Find the duty cycle range.

9. A buck–boost converter that is switching at 50 kHz is supplied with an input voltage that varies between 5 V and 10 V. The output is required to be regulated at 15 V. A load resistor of 15 Ω is connected across the output. If the maximum allowable inductor current ripple is 10% of the average inductor current, estimate the value of the inductance to be used in the buck–boost converter.

10. For Problem 9, if the output voltage ripple specification is 1% of the output voltage, estimate the output capacitor value.

11. For a buck converter supplying a 10 A load, the inductor current ripple is designed to be 10% of the maximum load current. Calculate the load current at which the converter is at the boundary of CCM and DCM.

12. A forward converter is switched at 50 kHz with a duty cycle of 0.3. It is supplying a 50 W load at an output voltage of 5 V. The input to the forward converter is derived from the 230 V mains by using a capacitor input filter rectifier. Estimate the turns ratio of the transformer.

13. For Problem 12, the peak magnetizing current is 20 mA. If the output inductor current ripple is 10% of the average inductor current, calculate the peak current requirement for the primary switch (IGBT).

14. For Problem 12, if the transformer core flux reset is done by using a diode–resistor connected across the primary, calculate the power dissipated in the flux reset resistor R_f if the transformer is ideal.

15. For Problem 13, if the voltage rating of the IGBT is 1200 V, calculate the range of the flux reset resistor R_f. If the value of R_f is selected at the midpoint of the obtained range, what is the voltage developed across the IGBT switch during the turn-OFF when (a) transformer is ideal and (b) transformer has leakage inductance? Is the developed voltage within the IGBT rating?

16. A forward converter is switched at 50 kHz with a duty cycle of 0.3. It is supplying a 50 W load at an output voltage of 5 V. The input to the forward converter is derived from the 230 V mains by using a capacitor input filter rectifier. A demagnetizing winding with the same number of turns as the primary winding is used to reset the core flux. What is the voltage developed across the primary switch during the OFF-state?

17. A push–pull converter is switched at 50 kHz with a duty cycle of 0.4. It is supplying a 48 W load. The turn ratio of the center-tapped transformer is 5. The input to the push–pull converter is a 12 V battery. The output inductor current ripple is 10% of the load current.
- **a.** What is the output voltage?
- **b.** What is the peak current through the primary switch?
- **c.** What is the voltage developed across the primary switch when it is OFF?

18. A half-bridge converter is switched at 50 kHz with a duty cycle of 0.4. It is supplying a 48 W load. The turn ratio of the center-tapped transformer is 5. The DC link of the half-bridge converter is from a 12 V battery. The output inductor value is 10 mH. The output inductor current ripple is 10% of the load current.
- **a.** What is the output voltage?
- **b.** What is the peak current through the primary switch?
- **c.** What is the voltage developed across the primary switch when it is OFF?
- **d.** What is the average current of the secondary diodes?
- **e.** What is the rms current of the secondary diodes?
- **f.** What is the peak inverse voltage of the secondary diodes?
- **g.** What is the range of value for the flux walking capacitor so that the DC voltage

developed across the capacitor does not exceed 10% of the DC-link voltage?

19. A full-bridge converter is switched at 50 kHz with a duty cycle of 0.4. It is supplying a 48 W load. The turn ratio of the center-tapped transformer is 5. The DC link of the full-bridge converter is from a 12 V battery. The output inductor value is 10 mH. The output inductor current ripple is 10% of the load current.
- **a.** What is the output voltage?
- **b.** What is the peak current through the primary switch?
- **c.** What is the voltage developed across the primary switch when it is OFF?
- **d.** What is the average current of the secondary diodes?
- **e.** What is the rms current of the secondary diodes?
- **f.** What is the peak inverse voltage of the secondary diodes?
- **g.** What is the range of value for the flux walking capacitor so that the DC voltage developed across the capacitor does not exceed 10% of the DC-link voltage?

20. A flyback converter is switched at 50 kHz with a duty cycle of 0.4. It is supplying a 48 W load. The turn ratio is 5. The input voltage of the flyback converter is derived from a 12 V battery. The flyback inductor value as viewed from the primary side is 10 mH.
- **a.** What is the output voltage?
- **b.** What is the peak current through the primary switch?
- **c.** What is the peak current through the secondary diode?
- **d.** What is the voltage developed across the primary switch when it is OFF?
- **e.** What is the peak inverse voltage of the secondary diode?

ANSWERS

Fill in the Blanks

1. SPDT
2. filter
3. two
4. two
5. discontinuous
6. voltage
7. smaller
8. load; capacitor
9. amp-second or charge balance
10. charging
11. discharging
12. larger
13. I_{in}; input; Δi_L; ripple
14. load; capacitor
15. smaller; larger
16. average; AC ripple
17. load; capacitor
18. ideal
19. inductor current
20. continuous conduction mode
21. discontinuous conduction mode
22. invariant
23. continuous conduction mode
24. discontinuous conduction mode
25. buck; chopper; isolation
26. input chopper
27. linearly
28. bifilar
29. 0.5
30. positive; unidirectional.
31. positive and negative
32. half
33. half
34. DC link
35. DC-link voltage
36. DC link
37. half
38. less than
39. buck–boost
40. least
41. infinite
42. zero
43. quasi

CHAPTER 6

DC–AC Switched-Mode Converters

Learning Objectives

After reading this chapter, you will be able to:

- understand the different types of inverters.
- distinguish between square wave and PWM operation and understand the effect of PWM on the output harmonics.
- obtain the Fourier series for a PWM wave pattern.
- understand the operation of various inverter schemes.

DC–AC switched-mode converters are circuit configurations that convert DC into AC. They are popularly called inverters. The inverters are used in many power electronic applications such as AC motor controls, pumping applications, uninterruptible power supplies, electronic ballasts, electric vehicles, frequency converters, induction heating and so on. An inverter is a circuit or system that delivers AC power when energized from a source of DC power. Alternately, one may say that inverters are DC-to-AC converters providing the reciprocal function of rectification. Rectifiers change AC into DC whereas the inverters behave in exactly the opposite manner by converting DC into AC.

DC–AC converters appear in many applications with different nomenclatures. However, they can be classified into three broad categories, namely,

1. choppers;
2. oscillators;
3. inverters.

As discussed in the previous chapter, the choppers convert the input DC voltage to pulsed output. The outputs of the choppers have a non-zero average. The terms oscillators and inverters have been used interchangeably in the literature. The distinction between the two categories is somewhat arbitrary. The distinction is more in connotation as there is no clear line of demarcation between oscillators and inverters. In general, the term inverter is used where the emphasis is to provide AC power for some other circuit or equipment. Here the focus is on parameters like efficiency, control, protections, etc. A circuit where the emphasis is on producing an AC output as a reference signal, wherein the focus is on frequency stability and drift, has traditionally been called an oscillator even though the operational principles for the oscillator and the inverters are similar.

An inverter is an electronic circuit which converts a DC voltage source to an AC voltage source of specified magnitude and frequency. The AC output waveform may be a square wave, quasi-square wave or sine wave. The primary component in the inverter is the power semiconductor switch. The number of switches can vary from one switch for a self-oscillating-type inverter to a very large number of switches for a multi-level inverter. Based on the strategy used for driving the power semiconductor switches ON and OFF, the inverters are broadly classified as

1. self-oscillating type;
2. driven type.

In the self-oscillating-type inverters the power switching elements are turned ON and OFF by the inherent non-linear characteristics of the magnetic and the semiconductor devices. However, this type of inverter cannot be easily controlled. In the driven type of inverter the switching elements are turned ON and OFF by external control circuits. The outputs of the switched-driven inverters are amenable for control. In the sections to follow, the switching inverters belonging to both these categories are discussed.

6.1 Inverter Topologies

The inverter is composed of only power semiconductor switches and their drive circuits. Any of the controlled power semiconductor switches discussed in Chapter 1 can be used. Until the advent of the fully controlled switches like the power bipolar junction transistors (BJTs), metal oxide semiconductor field effect transistors (MOSFETs) and insulated gate bipolar transistors (IGBTs), the inverters primarily consisted of the thyristor-based switches. However, now the MOSFETs and IGBTs are more popular in the inverter topologies due to control on both the turn-OFF and turn-ON of the device coupled with better switch transition characteristics. The drive circuits for the inverter switches are based on the discussion of Chapter 2. Topologically, the inverters can be classified into the single-phase inverters and the three-phase inverters depending on whether they are feeding a single-phase load or a three-phase load. Both these classes of inverter topologies will be discussed in the following sub-sections.

Single-Phase Topologies

Basic Two-Level Structure

The simplest inverter topology is composed of a single pole double throw (SPDT) switch. The self-oscillating types of inverters use this single SPDT switch configuration. The driven types, however, range from the simplest single SPDT configuration to more complex multiple throw switch configurations. Figure 6.1(a) shows the schematic of the single SPDT configuration inverter. The voltage across the load at any given instant is dependent on the switch position.

V_1, V_2 and V_3 are derived from DC sources such that $V_2 < V_3 < V_1$. If the pole of the SPDT switch is connected to V_1, then the output voltage V_o is $V_1 - V_3$; if the pole of the SPDT switch is connected to V_2, then the output voltage V_o is $V_2 - V_3$. There are only two levels that the output can assume and therefore this is a *two-level* inverter. The topology of Figure 6.1(a) is a generic configuration. A specific configuration using a single DC input source V_{dc} is shown in Figure 6.1(b). Here, $V_1 = V_{dc}$, and V_2 and V_3 are connected to the circuit ground. The output voltage is measured with respect to the circuit ground. The output voltage V_o can assume either V_{dc} or 0 V depending on the switch pole position. Such a two-level inverter wherein one of the output voltage levels is 0 V is called a *chopper*. As the output voltage swings between 0 and V_{dc}, it contains a finite non-zero average value.

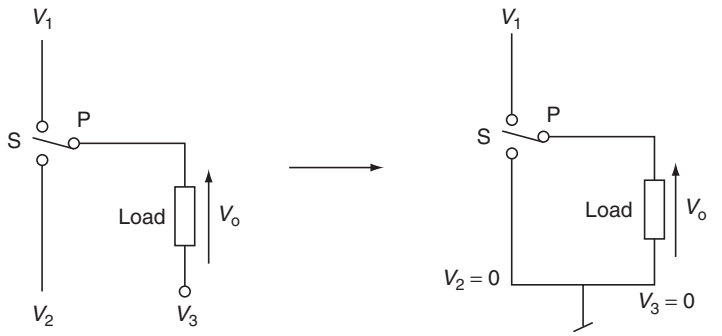


Figure 6.1 Simple single SPDT switch topology: (a) Generic configuration; (b) inverter configuration with V_2 and V_3 connected to the circuit ground.

Figure 6.2(a) is the implementation schematic of the two-level chopper topology depicted in Figure 6.1(b). S1 and S2 are the two power semiconductor switches used to implement the SPDT switch. When S1 is ON, V_o is at V_1 and when S1 is OFF and S2 is ON, V_o is 0 V. For inductive loads, the internal body diode of the IGBT provides the inductive energy freewheeling path. When S1 is ON and S2 is OFF, the load current of an inductive load flows from the pole of the SPDT switch to the ground. When S1 is turned OFF and S2 is turned ON, the inductive load current freewheels through the body diode of S2 till it reduces to zero.

Figure 6.2(b) shows the implementation schematic of the generic two-level topology. One end of the load is connected to the pole "p" (midpoint) of S1 and S2 switch arm and the other end of the load is connected to the midpoint of the capacitor arm as shown in Figure 6.2(b). The midpoint of the capacitor arm is at potential V_3. The output voltage can assume either of the two levels $(V_{dc} - V_3)$ or $(-V_3)$. Usually the capacitor values for C_1 and C_2 are the same and therefore $V_3 = V_{dc}/2$. In such a case the two levels of the inverter output voltage are $V_{dc}/2$ and $-V_{dc}/2$.

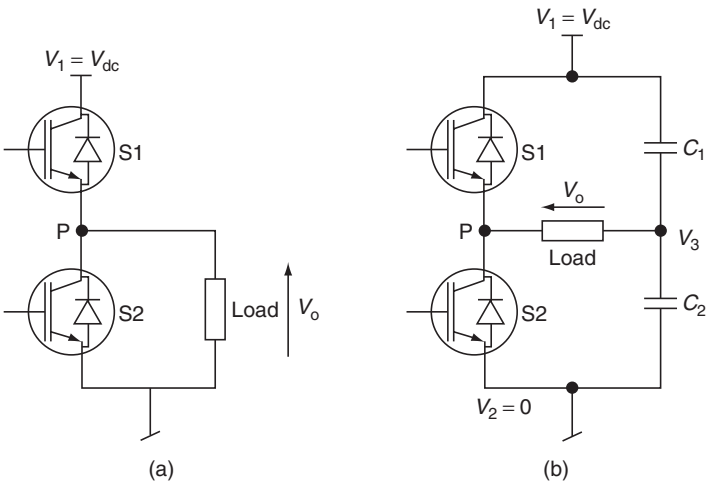


Figure 6.2 Implementation schematic of (a) the chopper; (b) a two-level inverter.

Multi-Level Structures

To have better control on the harmonic content of the output voltage waveform, it is desirable to have the output voltage that can assume many more levels than just two. In such cases, the simple two-level inverter topology will not suffice. One will have to use more complex configurations to achieve multi-level outputs. The number of switches for multi-level converters will increase depending on the number of levels required. Figure 6.3(a) gives the topology of a generic four-level inverter. The output voltage can assume four different levels depending on the switching pattern of the two SPDT switches. Table 6.1 gives the output levels in relation to the switching pattern.

Figure 6.3(b) gives an implementation schematic of Figure 6.3(a) wherein there is only one input DC voltage source. Here $V_1 = V_3 = V_{dc}$ and $V_2 = V_4 = 0$ V. Based on these constraints, Table 6.1 is modified as Table 6.2

From Table 6.2, it can be observed that with a single input DC voltage source, the four-level structure of Figure 6.3(a) reduces to a three-level inverter wherein the output voltage can assume only V_{dc}, 0 and $-V_{dc}$.

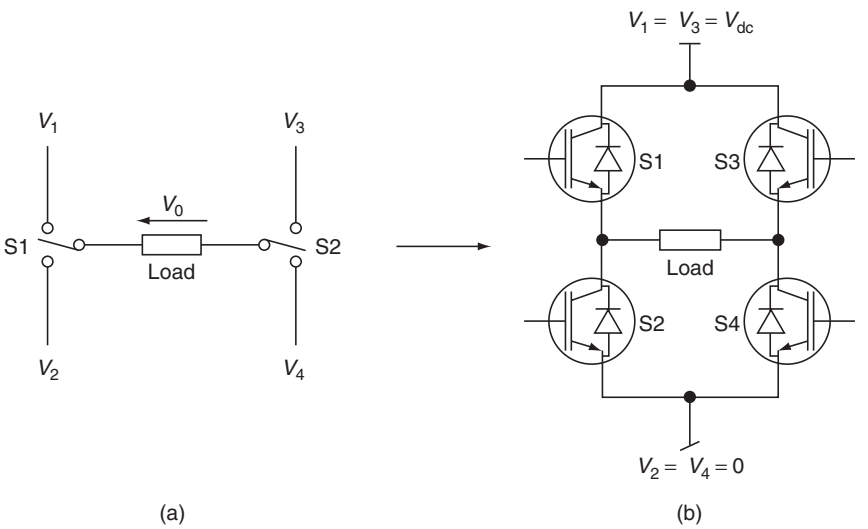


Figure 6.3 Generic four-level topology; (b) implementation schematic of a three-level structure by applying single input DC voltage source to the generic four-level topology.

Table 6.1 Output voltage levels in relation to the switching pattern

S1 connected to	S2 connected to	Output voltage
V_1	V_4	$V_1 - V_4$
V_1	V_3	$V_1 - V_3$
V_2	V_4	$V_2 - V_4$
V_2	V_3	$V_2 - V_3$

Table 6.2 | Output voltage levels with single input DC voltage source

S1 connected to	S2 connected to	Output voltage
V_{dc}	0	V_{dc}
V_{dc}	V_{dc}	0
0	0	0
0	V_{dc}	$-V_{dc}$

The schematic of Figure 6.4(a) shows a single pole triple throw (SPTT) switch S wherein the pole P is connected to one end of the load and the throws are connected to DC voltage sources V_1, V_2 and V_3. The other end of the load is connected to another DC voltage source V_4. Figure 6.4(b) shows the schematic wherein the SPTT switch is implemented as two SPDT switches. This forms a *three-level arm*.

The output voltage can assume three different levels depending on the switching pattern of the two SPDT switches. Table 6.3 gives the output levels in relation to the switching pattern.

The three-level arm can be applied on either side of the load as shown in Figure 6.5(a). Figure 6.5(b) shows the implementation with only SPDT switches. This forms the *three-level arm bridge* topology.

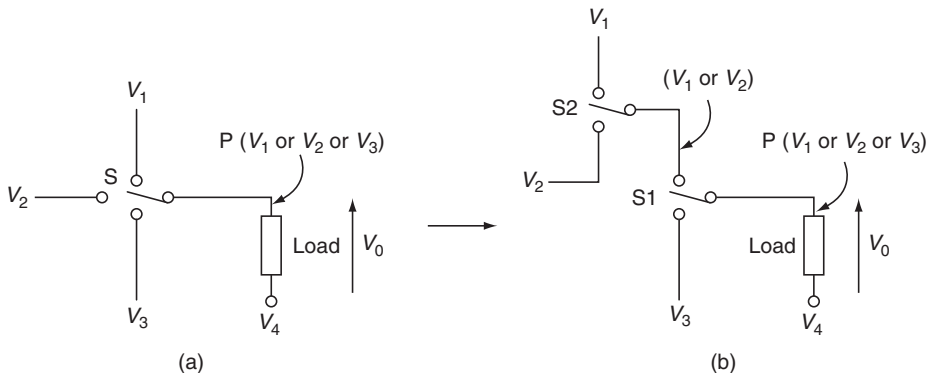


Figure 6.4 | Three-level arm: (a) SPTT arm; (b) SPTT switch replaced by two SPDT switches.

Table 6.3 | Output voltage levels in relation to the switching pattern for schematic of Figure 6.4

S1 connected to	S2 connected to	Output voltage
Pole of S2	V_1	$V_1 - V_4$
Pole of S2	V_2	$V_2 - V_4$
V_3	V_1 or V_2	$V_3 - V_4$

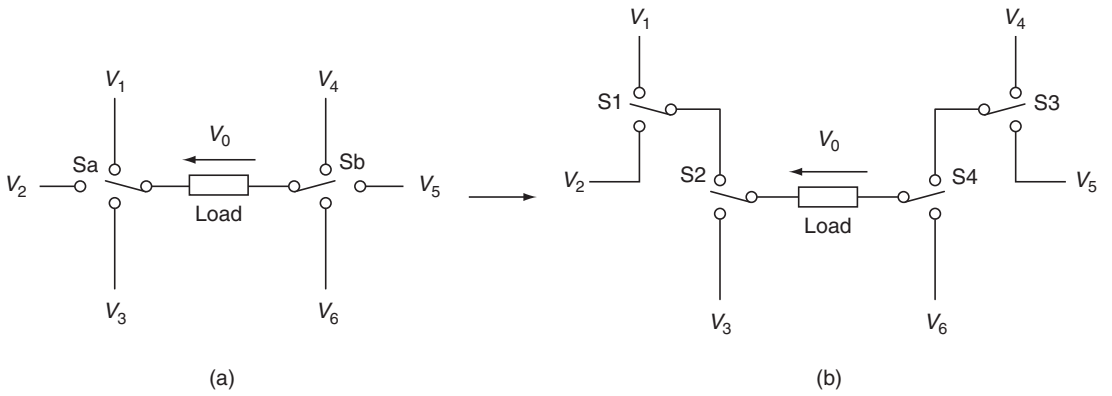


Figure 6.5 | Single-phase three-level arm bridge: (a) Schematic; (b) SPTT switches replaced by SPDT switches.

The output voltage can assume nine different levels depending on the switching pattern of the SPDT switches. Table 6.4 gives the output levels in relation to the switching pattern.

To maintain symmetry if $V_1 = V_4$, $V_2 = V_5$ and $V_3 = V_6$, from Table 6.4 we have $V_1 - V_4$, $V_2 - V_5$ and $V_3 - V_6$ are zero voltage levels and therefore the inverter output can assume only seven levels. Figure 6.6 shows the circuit implementation of the multi-level inverter topology described in Figure 6.5(b).

The three-level arm can be extended to an n-level arm and applied to both sides of a load to form an n-level arm bridge. This two-arm n-level bridge inverter is depicted in Figure 6.7. In general, for a bridge arm,

Table 6.4 | Output voltage levels in relation to the switching pattern for schematic of Figure 6.5

S1 connected to	S2 connected to	S3 connected to	S4 connected to	Output voltage
V_1	Pole of S1	V_4	Pole of S3	$V_1 - V_4$
V_2	Pole of S1	V_4	Pole of S3	$V_2 - V_4$
V_1	Pole of S1	V_5	Pole of S3	$V_1 - V_5$
V_2	Pole of S1	V_5	Pole of S3	$V_2 - V_5$
V_1 or V_2	V_3	V_4	Pole of S3	$V_3 - V_4$
V_1 or V_2	V_3	V_5	Pole of S3	$V_3 - V_5$
V_1	Pole of S1	V_4 or V_5	V_6	$V_1 - V_6$
V_2	Pole of S1	V_4 or V_5	V_6	$V_2 - V_6$
V_1 or V_2	V_3	V_4 or V_5	V_6	$V_3 - V_6$

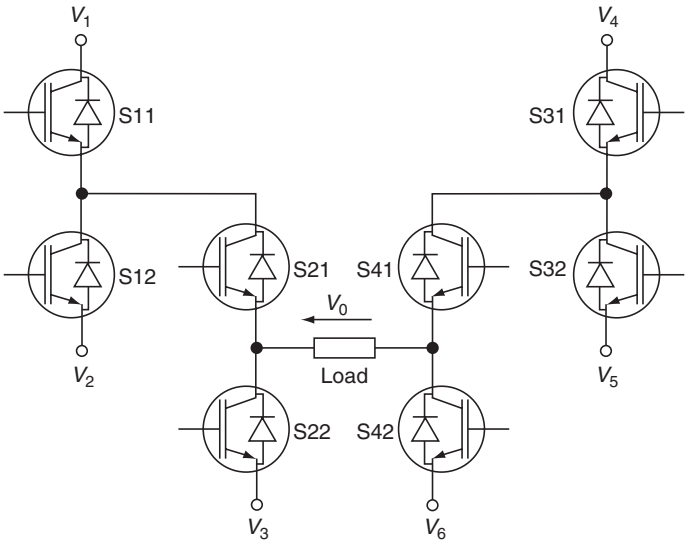


Figure 6.6 | Circuit implementation of the multi-level inverter topology of Figure 6.5(b).

if there are *n*-input DC sources then there are *n* possible levels that the arm pole can assume. The *n* sources are connected to the throws of a single pole *n*-throw switch. To build an equivalent single pole *n*-throw switch, $n-1$ SPDT switches are needed.

For the two-arm *n*-level bridge inverter topology as shown in Figure 6.7, the maximum number of possible levels for each arm is *n*. The maximum number of levels that the output can assume is $n \cdot n = n^2$. It would be n^2 only if all the voltages on both sides of the load are unique. However, if the arms are symmetric in voltages, then the combinations wherein the same *n* voltages are applied to both bridge arms in a symmetrical manner will lead to zero load voltage. Under this condition, there will be *n* possibilities wherein the voltages across the load is zero. Thus, out of the possible n^2 possibilities, *n* possibilities are removed and combine to form a zero voltage level. Therefore, the number of possible output voltage levels in this case is $n^2 - n + 1$.

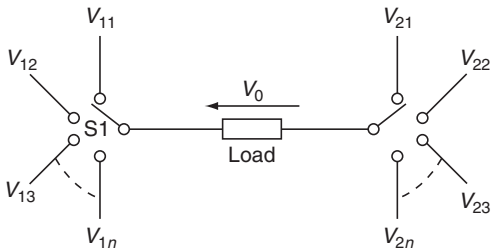


Figure 6.7 | Two-arm *n*-level arm bridge inverter topology.

If "m" is the required number of output voltage levels, then

$$m = n^2 - n + 1$$

Therefore the number of distinct voltage sources required is

$$n = \frac{1 \pm \sqrt{4m - 3}}{2} \tag{6.1}$$

The number of distinct voltage sources should be an integral number; therefore, the solution of Eq. (6.1) should be a positive integer. If both the solutions of the quadratic are not appropriate then the required number of levels "m" specified is inappropriate.

Three-Phase Topologies

The previous sub-section discussed the inverter topologies for a single-phase load that is supplied, in general, by a two-arm bridge wherein the single-phase load is connected in between the two arms of the inverter. In the case of a three-phase inverter, the single-phase inverter is extended to three-phase loads by including one more bridge arm. Thus, three-phase inverter topologies consist of three bridge arms wherein each arm can be of arbitrary level. However, practical inverters consists of three bridge arms wherein all the bridge arms are identical and the throws of the single pole multi-throw switches are supplied by the same voltage sources across all three bridge arms. Figure 6.8(a) shows a two-level three-arm bridge converter supplying a three-phase load. Figure 6.8(b) shows the circuit implementation. The three-arm bridge can either supply a three-phase star-connected load as shown in Figure 6.9(a) or a three-phase delta-connected load as shown in Figure 6.9(b). Instead of a two-level arm, one may synthesize a multi-level three-phase inverter by using a multi-level structure for the arms of the inverter along lines discussed for the single-phase inverter topologies.

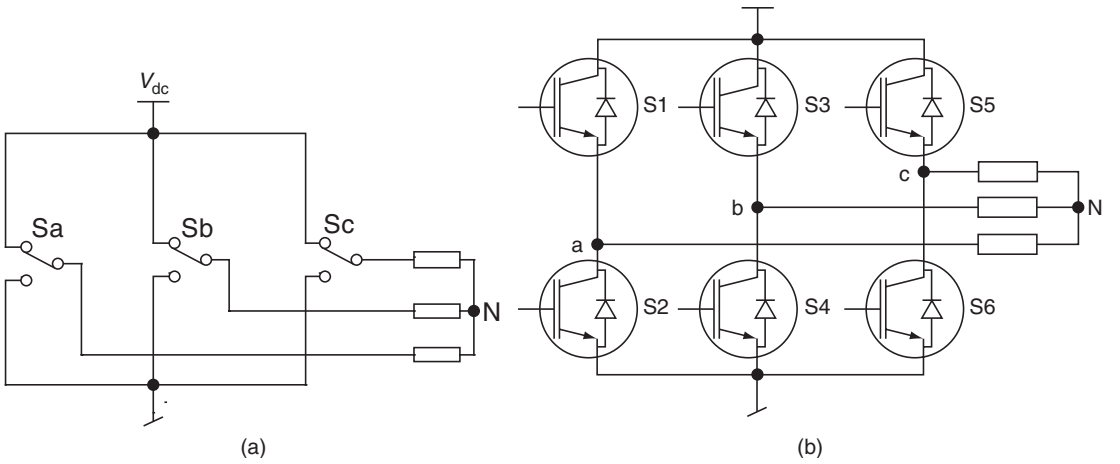


Figure 6.8 (a) Inverter with three-arm bridge; (b) implementation schematic of the three-arm bridge topology.

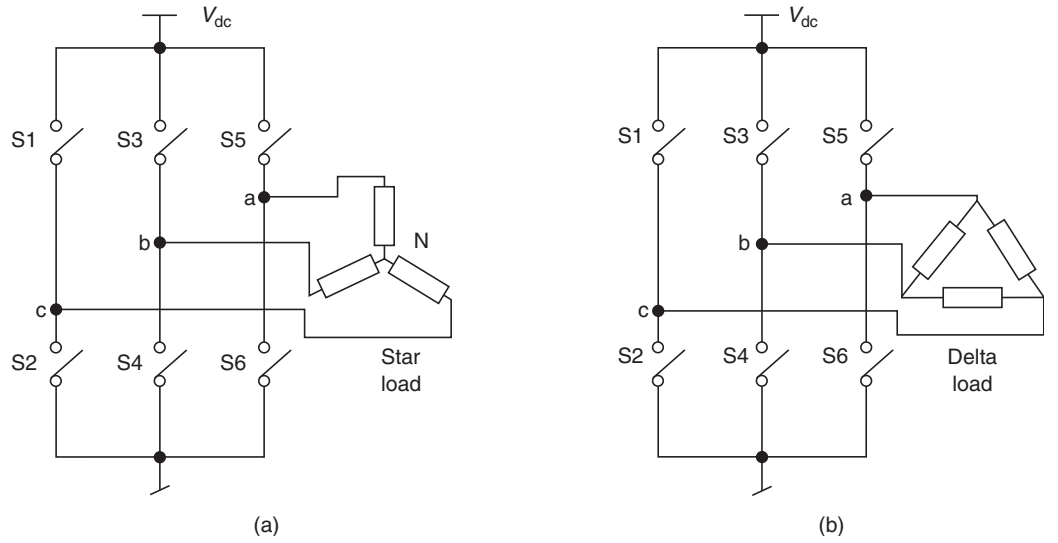


Figure 6.9 | Three-arm bridge connected to (a) three-phase star load; (b) three-phase delta load.

6.2 Self-Driven Inverters

The self-driven inverters are a class of DC–AC converters wherein the power semiconductor devices are operated without externally applied control signals to the bases or gates of the devices. Based on the operation mode of the power semiconductor devices, they can be broadly classified as follows:

1. **Power Semiconductor Devices Operated in the Linear Region:** These types are used as oscillators for reference generation, linear audio applications, etc. wherein the signal power is very low.
2. **Power Semiconductor Devices Operated Alternatively in Saturation and Cut-Off Regions:** These types are called switching self-driven inverters. They are used for supplying power to power supplies that need voltage translation.

In this section, only the switching type of self-driven inverters will be discussed, wherein the power semiconductor switches are driven to saturation and cut-off alternately.

Generally, in inverters the transformers are operated in the linear magnetic region. The saturation of the magnetic core is usually avoided because the operation in this region greatly increases the hysteresis loss component that deteriorates the efficiency. Though the transformer driven into magnetic saturation becomes a source of power dissipation, the combination of a transformer driven to saturation and power transistors can produce an inverter with reasonably high efficiency as well as high reliability and low component count. Core saturation is the mechanism responsible for switching and timing in these types of inverters.

The important features of saturable-core inverters are:

1. Switching transitions occur due to the characteristics of the magnetic core material and the static characteristics of the power transistors. Core saturation alone cannot account for the operation, nor can only the transfer characteristics of the power transistor.

2. Despite positive feedback, the oscillation is not of the LC type.
3. The frequency of the oscillation of such an inverter is dependent on the magnetic characteristics of the core material and is independent of transistor parameters and temperature in an ideal circuit. The ideal circuit also exhibits zero regulation. However, in a practical circuit, the frequency of oscillation is dependent on the transistor parameters and temperature too.
4. The ideal saturable-core inverter generates a frequency that is directly proportional to the input DC voltage. The wave shape is a symmetrical rectangular or square wave with a 50% duty cycle.

Saturable-Core Inverter

The circuit schematic shown in Figure 6.10 is a self-driven inverter wherein the oscillations are induced due to the non-linear characteristics of the magnetic core.

Circuit Operation

When V_{in} is applied at switch ON of the inverter, current flows through R either through the path N_{b2}–R_2–Q_2 (base–emitter) or through the path N_{b1}–R_1–Q_1 (base–emitter). The path which is decided at start-up depends on the characteristics of Q_1 and Q_2. It should be noted that Q_1 and Q_2 will never have identical characteristics.

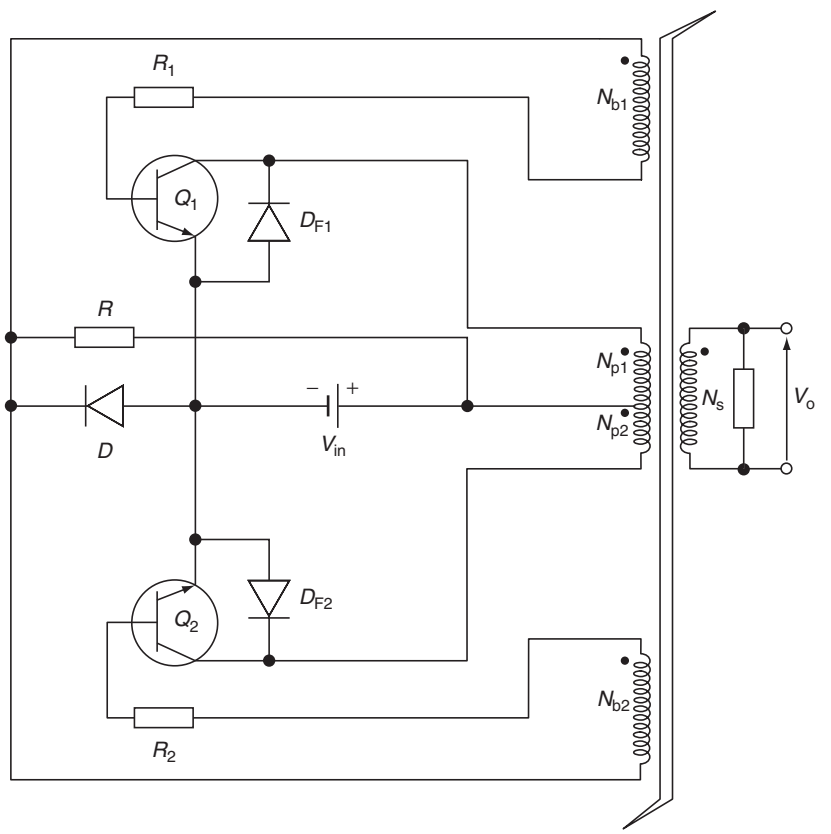


Figure 6.10 | Saturable-core inverter topology.

Let Q_2 have characteristics such that its V_{ce} is slightly lower than that of Q_1 for a given base current. Then at switch ON, current from V_{in} flows through R and N_{b1}–R_1–Q_1 (base–emitter) and N_{b2}–R_2–Q_2 (base–emitter). If $N_{b1} = N_{b2}$ and $R_1 = R_2$, at start both transistors get equal base currents. However due to the non-identical characteristics of Q_1 and Q_2 as stated earlier, V_{ce2} will be lesser than V_{ce1}. The voltage across N_{p1} is $V_{in} - V_{ce1}$ which tries to make the non-dot end positive. The voltage across N_{p2} is $V_{in} - V_{ce2}$ and it tries to make the dot end positive. As V_{ce2} is lesser than V_{ce1}, the dot end of the windings will tend to become more positive. As a consequence, from the depicted "dot" polarity, N_{b2} "dot" is more positive compared to N_{b1} non-dot end. This tends to turn Q_2 more towards conduction and Q_1 more towards the OFF-state. This re-generatively builds up until Q_2 is in saturation and Q_1 is in cut-off state. At this point the base current takes the path N_{b2}–R_2–Q_2 (base–emitter)–D.

When Q_2 is in saturation, V_{p2} (the voltage across N_{p2} is a constant) of magnitude V_{in} (neglecting the saturation drops across the collector–emitter of the transistors), the current through the winding N_{p2}, that is, i_{c2}, keeps increasing linearly as shown in Figure 6.11 because

$$V_{in} = L_{p2} \frac{di_{c2}}{dt}$$

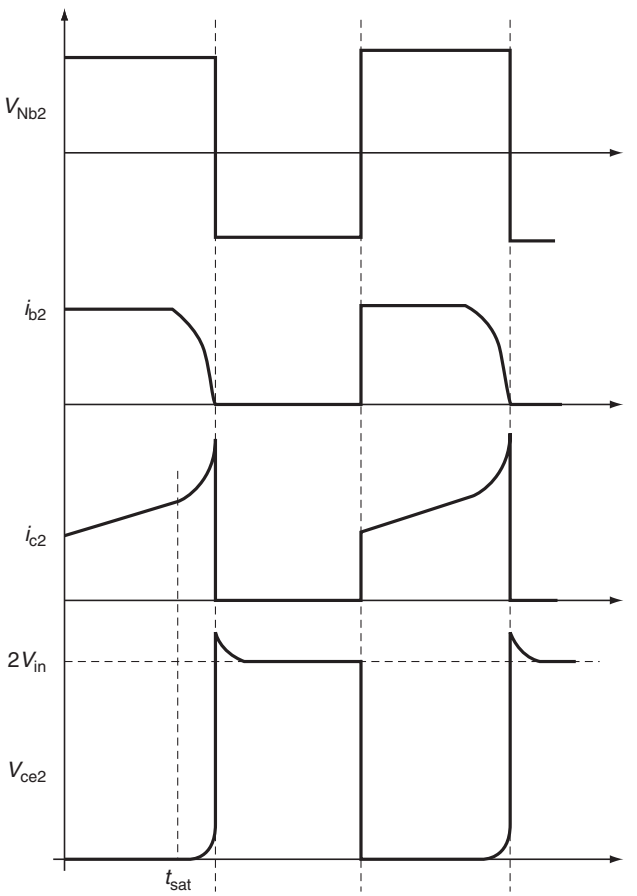


Figure 6.11 | Waveforms for transistor Q_2.

where L_{p2} is the inductance of winding N_{p2}.

As the collector current increases, the transformer begins to saturate at time t_{sat}. As a result the current i_{c2} starts to shoot up as indicated in Figure 6.11 due to reduction in permeability. The induced voltage across N_{b2} reduces as the transformer core goes into saturation, thereby decreasing the base drive for Q_2. This causes Q_2 to come out of saturation region. A reduction in induced voltage at the dot end of the winding implies an equivalent increase at the non-dot end of the base windings. The N_{b1} polarity is such that it is 180° out-of-phase with respect to N_{b2}. Therefore Q_1 tries to cut-in. This process re-generatively builds up until Q_1 goes to saturation and Q_2 goes OFF.

The above process is repeated with transistor Q_1 also and this continues indefinitely, thereby giving rise to oscillations. The waveforms for Q_1 are similar to that of Q_2 as shown in Figure 6.11. However, Q_1 waveforms are 180° out-of-phase with respect to the waveforms of Q_2.

Oscillation Frequency

Much insight into the behavior of the saturable-core inverter can be gained from the basic equation defining its frequency of oscillation. From the theory on magnetics (discussed in Chapter 7), it can be shown that

$$E = 4K_f NA_c B_m f \tag{6.2}$$

where E is the root mean square (rms) value of the induced voltage; K_f the form factor; f the frequency in Hz; N the number of turns in the winding; A_c the core cross-section area in m^2; B_m the maximum operating flux density in Tesla. Rearranging Eq. (6.2), the oscillation frequency can be determined from

$$f = \frac{E}{4K_f NA_c B_m} \tag{6.3}$$

For this particular application of saturable-core inverter, the waveform is a square waveform and therefore $K_f = 1$ and $E = V_{in}$. Further, B_m (the maximum flux density) has to be chosen equal to the material saturation flux density B_{sat} because it is required that the core saturates every switching cycle. Therefore $B_m = B_{sat}$ for these inverters.

Design Equations

In designing the saturable-core inverter, first determine the transistor ratings and select a suitable transistor:

$$I_{cm} > I_{in} \tag{6.4}$$

where

$$I_{in} = P_{in}/V_{in};\ P_{in} = P_o/\eta$$

η is the efficiency of the transformer – 80% efficiency is a conservative value that can be used for design; I_{cm} is the continuous maximum rating of the transistor. Referring to the waveforms shown in Figure 6.11, the voltage rating for the transistor is given as

$$V_{ceo} > 2V_{in} \tag{6.5}$$

A transistor should be chosen such that the inequalities in Eqs. (6.4) and (6.5) are satisfied. After having chosen the transistor, decide on the turns ratio for the base–primary winding such that,

$$n_b = \frac{V_b}{V_{in}} = \frac{N_{b2}}{N_{p2}} = \frac{N_{b1}}{N_{p1}} \tag{6.6}$$

where V_b is the voltage across the base winding (taken to be around 3 V). The transformer is designed using the area product technique that will be discussed in Chapter 7. The area product for this transformer is derived in Appendix I. It is given as follows:

$$A_p = \frac{P_o}{4K_w J B_m f_s}\left(\frac{\sqrt{2}}{\eta} + \frac{\sqrt{2}n_b}{\eta h_{FE\,min}} + 1\right) \tag{6.7}$$

where K_w is the window factor = 0.4; J the current density which is typically 3×10^6 A/m^2
$B_m = B_{sat}$ of the material; f_s the oscillation frequency; P_o the output power; η the efficiency of the inverter which is usually around 0.8. The number of turns is given as follows [refer to Eq. (6.2)]:

$$N_{p1} = \frac{V_{inmax}}{4B_m A_c f_s} \tag{6.8}$$

$$N_{p2} = N_{p1}$$

$$N_s = \left(\frac{V_o}{V_{inmin} - V_{cesat}}\right) N_{p1} \tag{6.9}$$

where V_o is the secondary voltage and

$$N_{b1} = N_{b2} = n_b N_{p1} \tag{6.10}$$

The resistors R_1 and R_2 can be calculated as follows:

$$R_1 = R_2 = \frac{V_b - V_{besat} - V_D}{I_b} \tag{6.11}$$

where

$$I_b = \frac{I_{in}}{h_{FEmin}}$$

and dissipation in R_1 and R_2 are

$$P_{R1} = P_{R2} = I_{brms}{}^2 R_1 \tag{6.12}$$

where

$$I_{\text{brms}} = \frac{I_{\text{b}}}{\sqrt{2}}$$

The resistor R is required only at the time of starting the oscillations, when it is required to just cut-in one of the transistors till the regenerative action takes over. The value of R is given by

$$R = \frac{V_{\text{inmin}} - V_{\text{be-cutin}}}{I_{\text{b-cutin}}} \tag{6.13}$$

where

$$I_{\text{b-cutin}} = 0.2\left(\frac{I_{\text{in}}}{h_{\text{FEmin}}}\right)$$

and the power dissipation is given by

$$P_{\text{R}} = \frac{(V_{\text{inmax}} + V_{\text{D}})^2}{R} \tag{6.14}$$

Saturable Base Drive Inverter

In the previous section, the output transformer is designed such that it goes into saturation region and, hence compared to a transformer which operates in the linear region, the hysteresis loss is much greater. Further the power switches have to handle high peak currents at the time of core saturation. To avoid the short comings, the output transformer is designed to operate in the linear region, but another transformer is used for base drives of the power switches, as indicated in Figure 6.12, which is designed to go into saturation region. As a

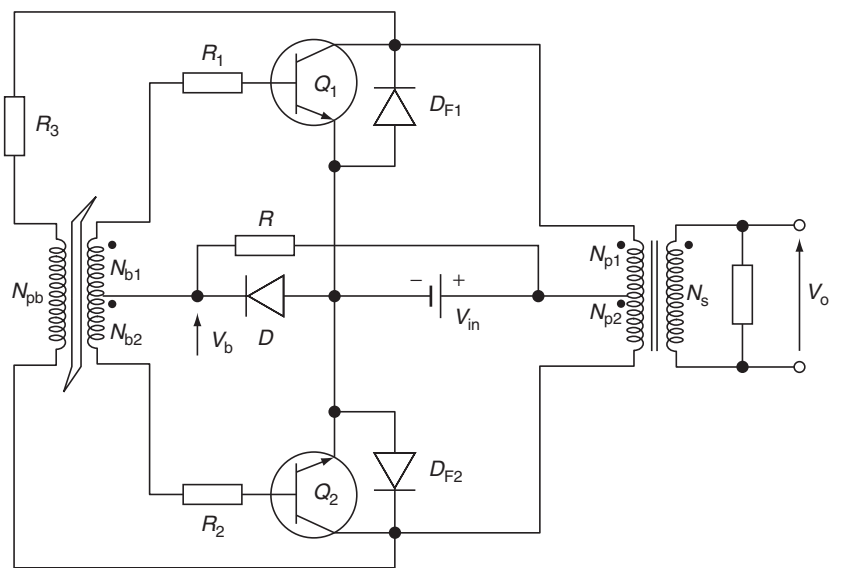


Figure 6.12 | Self-driven inverter with saturable base drive transformer.

consequence, better efficiency is obtained using this two-transformer technique. The operation is exactly similar to the single transformer saturable-core transformer inverter, except that for the base drive, a separate transformer is used which is saturable and that for the power output the transformer is designed to operate in the linear region.

The output transformer is designed as a push–pull transformer and the base drive transformer is designed along similar lines except that B_m is taken equal to B_{sat} of the core material. Here,

$$R_1 = R_2 = \frac{V_b - V_{be} - V_D}{I_b} \tag{6.15}$$

where V_b is the voltage developed across N_{b1} and N_{b2} of the saturable-core base drive transformer. The collector current is given as

$$I_c = I_{in} = \frac{P_o}{\eta V_{in}} \tag{6.16}$$

and the base current I_b is given as

$$I_b = \frac{I_c}{h_{FEmin}} \tag{6.17}$$

The following points should be noted while designing the base drive transformer.

1. The VA capacity is given as $P_T = V_b I_b$.

2. It is usual to set V_{Npb}, that is, voltage across primary of the base drive transformer, to V_{in}. As Q_1 and Q_2 collector–emitter voltages swing between 0 and $2V_{in}$, the resistor R_3 also drops a voltage of V_{in} across itself.

3. N_{pb} is calculated using the Faraday's Law given by

$$N_{pb} = \frac{V_{in}}{4B_m A_c f_s} \tag{6.18}$$

4. The secondary turns of the base drive transformer are given by

$$N_{b1} = N_{b2} = \left(\frac{V_b}{V_{in}}\right) N_{pb} \tag{6.19}$$

The value of V_b can be chosen to be around 3 V.

5. The primary current of the base drive transformer is given as

$$I_{Npb} = \frac{P_T}{V_{in}} \tag{6.20}$$

and

$$R_3 = \frac{V_{in}}{I_{Npb}} = \frac{V_{in}^2}{P_T} \tag{6.21}$$

At higher input voltages, N_{pb} may be large. The turns may be reduced in the base drive transformer by adding a low-voltage winding in the output transformer as shown in Figure 6.13.

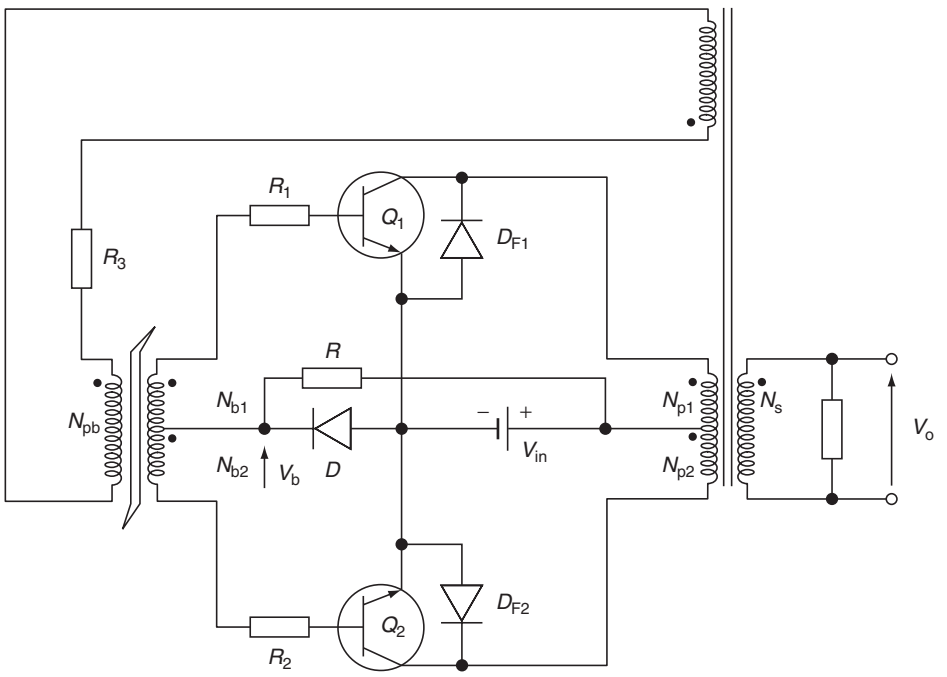


Figure 6.13 | Low-voltage base drive transformer with additional output transformer winding.

6.3 Driven Inverter

In the case of driven inverters, the power switches are driven through external control circuits where the ON-time and the OFF-time are controllable. Because of the controllable nature of the power switch states, any desired output wave shape, amplitude and frequency can be easily achieved. There are various control strategies through which one could achieve either a quasi-square wave output or a waveshape with specific harmonic minimization or even a sinusoidal waveshape.

For any control strategy, the power circuit configuration used is generally one of the topologies discussed in Section 6.1. The following three configurations are commonly used for the driven inverter circuits:

1. the push–pull configuration;
2. the half-bridge configuration;
3. the full-bridge configuration.

Push–Pull Configuration

The push–pull configuration deviates from the other configurations in that the load is connected across the throws of the SPDT switches rather than at the pole. The schematic is shown in Figure 6.14(a). It consists of a SPDT switch with split load as shown. The split load is connected across the throws of the switch S. The input voltage source V_i is connected between the split load and the pole of the switch S. Figure 6.14(b) gives the schematic of the push–pull topology wherein the load is replaced by a transformer with the actual load

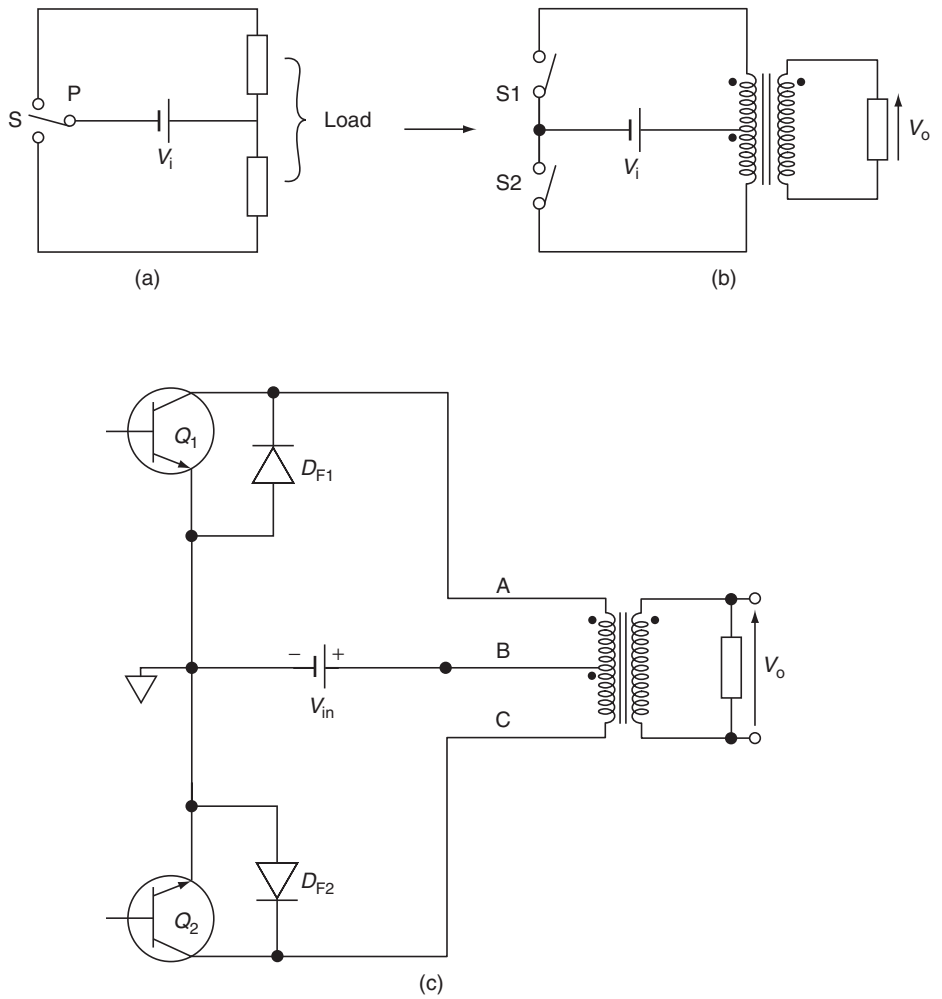


Figure 6.14 (a) Generic push–pull configuration with split load; (b) push–pull configuration with transformer load; (c) circuit implementation of the push–pull configuration with transformer load.

connected in the secondary of the push–pull transformer. The SPDT switch is replaced by two single pole single throw (SPST) switches S1 and S2. Figure 6.14(c) gives the circuit implementation of the push–pull DC–AC inverter wherein the switches S1 and S2 of Figure 6.14(b) are replaced by power semiconductor switches Q_1 and Q_2, respectively. The switches Q_1 and Q_2 can be BJTs or MOSFETs or IGBTs.

Consider the circuit shown in Figure 6.14(c) wherein the load is inductive. Let Q_2 be in the ON condition. Then $V_{BC} = V_{in}$ and $V_{ce1} = V_{AC} = 2V_{in}$. If Q_2 is switched OFF, V_{BC} becomes negative because $\mathrm{d}i/\mathrm{d}t$ is negative as the current in the winding is cut-off and causes the voltage polarity reversal due to $L\mathrm{d}i/\mathrm{d}t$ effect. By magnetic induction, V_{BA} is now positive which acts against V_{in}, causing the magnetic energy stored in the inductor to freewheel through the path AB–V_{in}–D_{F1}–AB.

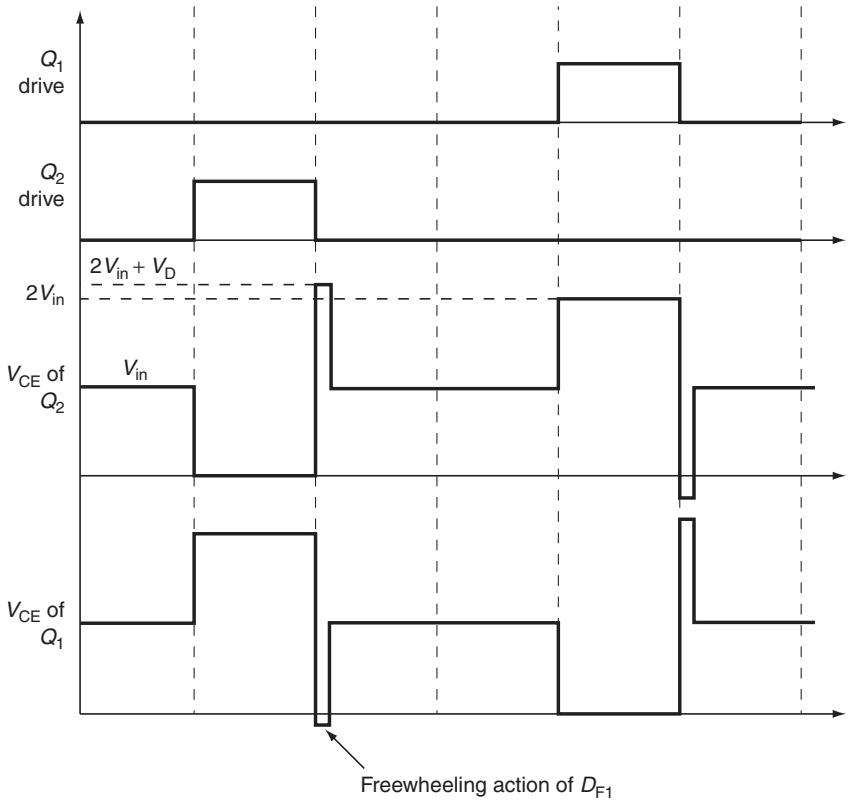


Figure 6.15 | Typical waveforms for the push–pull configuration.

Now when Q_1 is turned ON, $V_{BA} = V_{in}$ and $V_{ce2} = 2V_{in}$. On switch-OFF of Q_1, as discussed before, the magnetic energy freewheels through the path CB–V_{in}–D_{F2}–CB. Figure 6.15 illustrates the operation of the push–pull converter by depicting the typical waveforms.
The ratings for the power switching devices Q_1 and Q_2 are as follows:

$$V_{ceo} \text{ rating} > 2V_{in}$$

$$I_{cm} \text{ rating} > P_o/(\eta V_{in})$$

where P_o is the output power and η is the efficiency of the transformer.

Half-Bridge Configuration

The half-bridge inverter configuration is based on the generic two-level inverter topology of Figure 6.2. The implementation circuit of the half-bridge inverter is shown in Figure 6.16. Generally $C_1 = C_2$ and $R_1 = R_2$ in half-bridge inverter circuits. This means that the potential at point B is $V_{in}/2$.

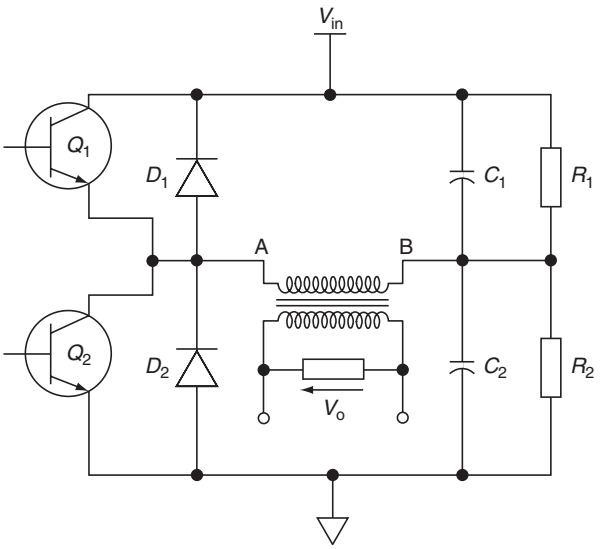


Figure 6.16 | Circuit schematic of the half-bridge inverter configuration.

Circuit Operation

Let Q_1 be ON to start with. The current flows from A to B in the primary of the load transformer. V_{AB} is positive. On switching OFF Q_1, V_{BA} becomes positive because the equivalent primary inductance present between A and B will cause point B to be more positive with respect to point A due to $L\mathrm{d}i/\mathrm{d}t$ effect. The inductive current freewheels through the path AB–C_2–D_2–AB, where the current is forced against the voltage across C_2 (i.e., $V_{in}/2$) and the path AB–C_1–V_{in}–D_2–AB, where the current is forced against $V_{in} - V_{C1}$ (i.e., $V_{in}/2$).

Similarly when Q_2 is ON, the current flows from B to A, V_{BA} is positive. On switching OFF Q_2, V_{AB} becomes positive because voltage across the equivalent primary inductance between the points A and B will reverse polarity according to Faraday's Law ($v = L\mathrm{d}i/\mathrm{d}t$) due to cutting off the current through the switch Q_2. The current freewheels through the paths BA–D_1–C_1–BA and BA–D_1–V_{in}–C_2–BA.

The power switching transistors should be selected such that

$$V_{ceo}\ \text{rating} > V_{in}$$

and

$$I_{cm}\ \text{rating} > P_o/(\eta V_{in}/2)$$

R_1 and R_2 are provided to divide the DC-link voltage V_{in} equally between the two capacitors C_1 and C_2. Generally, R_1 and R_2 are designed by allowing about 2% of the full load to flow through them. Therefore,

$$R_1 = R_2 = \frac{V_{in}/2}{0.02 I_{cm}}$$

Full-Bridge Configuration

The full-bridge inverter configuration is based on the generic four-level two-arm bridge topology. If there is only one input voltage source, then as discussed in Section 6.1, the full-bridge inverter configuration reduces to a three-level two-arm structure. The circuit implementation of the three-level two-arm full-bridge inverter is shown in Figure 6.17.

Circuit Operation

To start with, let Q_1 and Q_4 be turned ON. V_{AB} is positive and current flows from A to B in the primary of the load transformer. On switching OFF Q_1 and Q_4, V_{BA} becomes positive due to the sudden break in the current and consequent large negative di/dt that causes reversal of the voltage polarity across the primary of the transformer. This forces the current to freewheel through the path AB–D_3–V_{in}–D_2–AB against the DC-link voltage, V_{in}. Similarly, when Q_3 and Q_2 are turned ON, V_{BA} is positive and current flows from B to A. On switching OFF Q_3 and Q_2, V_{AB} becomes positive due to the inductance present between points A and B. This forces the current to freewheel through the path BA–D_1–V_{in}–D_4–BA against the DC-link voltage V_{in}. Here the transistors are selected such that

$$V_{ceo}\ \text{rated} > V_{in}$$

$$I_{cm}\ \text{rated} > P_o/(\eta V_{in})$$

It should be noted that in the case of a half-bridge configuration the voltage across points A and B (the central load arm) swings between $+V_{in}/2$ and $-V_{in}/2$, whereas in the case of a full-bridge configuration, the voltage across the points A and B swings between $+V_{in}$ and $-V_{in}$. Therefore for a given power P_o, the current rating required for the power devices in the half-bridge configuration is double that for the full-bridge configuration. Therefore, for higher powers, the full-bridge configuration is generally preferred.

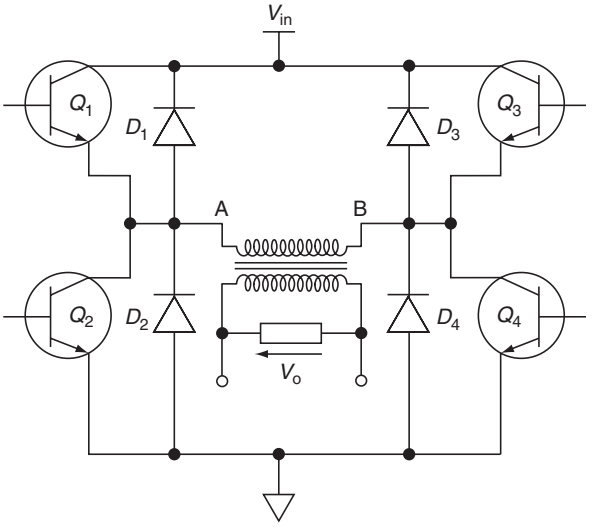


Figure 6.17 | Circuit schematic of the full-bridge inverter configuration.

6.4 Quasi-Square Wave Inverter

A waveform of the form shown in Figure 6.18 is called a "quasi-square wave". This is a waveshape that is easily achieved in a simple manner and hence it is popular in low-power, low-quality back-up power supplies. The quasi-square waveform has a large harmonic content and hence it is not suitable for transformer or motor loads as it would heat up the core material. However, it has its niche application as a back-up power supply for loads that have the rectifier–capacitor filter as the frontend.

An inverter which generates an output waveshape of the form shown in the Figure 6.18 is called a "quasi-square wave" inverter. Any of the basic configurations (push–pull, half-bridge or full-bridge) can be used with appropriate base drive signals for the power switches to generate the quasi-square wave. The parameter of the quasi-square wave is the dead band angle θ as indicated in Figure 6.18.

Circuit Operation

With reference to a full-bridge configuration like that shown in Figure 6.17, the drive signals for the power switches Q_1, Q_2, Q_3 and Q_4 to generate a quasi-square waveform are as shown in Figure 6.19. The switches Q_1 and Q_2 are switched 180° out-of-phase with respect to each other and Q_3 and Q_4 are similarly switched 180° out-of-phase with respect to each other. However the switching pattern for Q_3 and Q_4 are shifted in-phase by an angle α with respect to the switching pattern of Q_1 and Q_2. The angle α is twice the dead band angle θ. When Q_1 and Q_4 are ON, then V_{AB} is positive as shown in Figure 6.19. When Q_4 goes OFF, Q_3 would turn ON. During this time, Q_1 and Q_3 will be ON. This will ensure that the voltage V_{AB} is zero. If the load is inductive, the current will freewheel through Q_1 and D_3. When Q_3 and Q_2 are ON, V_{AB} is negative, as indicated. When Q_3 turns OFF, Q_4 will become ON. During this time, Q_2 and Q_4 are ON ensuring zero voltage across V_{AB}. The current now freewheels through Q_2 and D_4 if the load is inductive.

Output Regulation

To perform the output voltage regulation, the output voltage is sensed, rectified, filtered and processed. This filtered DC voltage is used for controlling the output voltage. The output quantity that can be controlled is one of the following:

1. output voltage flat top peak value;
2. output voltage half-cycle average value;
3. output voltage rms value.

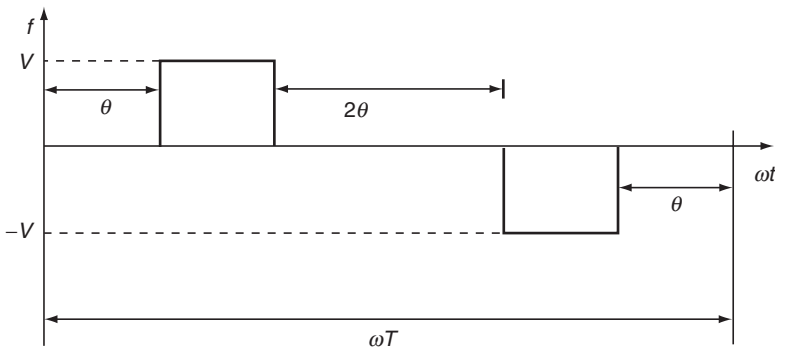


Figure 6.18 | Quasi-square waveform.

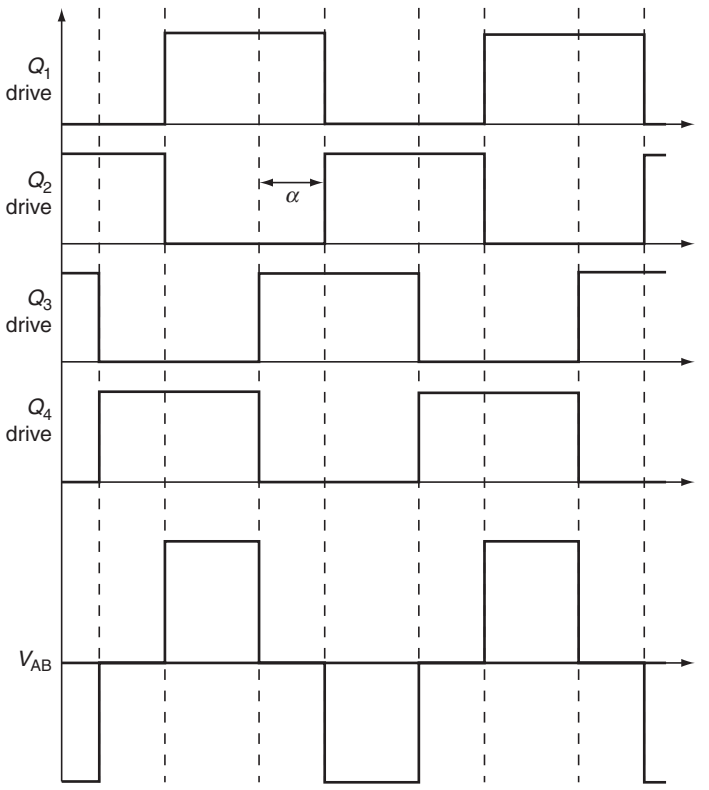


Figure 6.19 | Drive signals for quasi-square waveform generation.

If the duty ratio of the rectified output waveform is D, then

$$V_{\text{o-avg}} = VD$$

$$V_{\text{o-rms}} = V\sqrt{D}$$

$$V_{\text{o-avg}} = V_{\text{o-rms}}\sqrt{D}$$

Depending on the quantity to be controlled, the reference is the desired flat top peak value or the average value or the rms value. The rectified and filtered output waveform is accordingly processed depending on the control requirement.

The Fourier series representation for the quasi-square waveform shown in Figure 6.18 is given as

$$f(\omega t) = \frac{4}{\pi}\sum_{n=1}^{\infty}\left(\frac{V}{n}\right)\cos(n\theta)\sin(n\omega t) \tag{6.22}$$

where n is the harmonic number. It can be observed from Eq. (6.22) that the output voltage can be controlled in two ways:

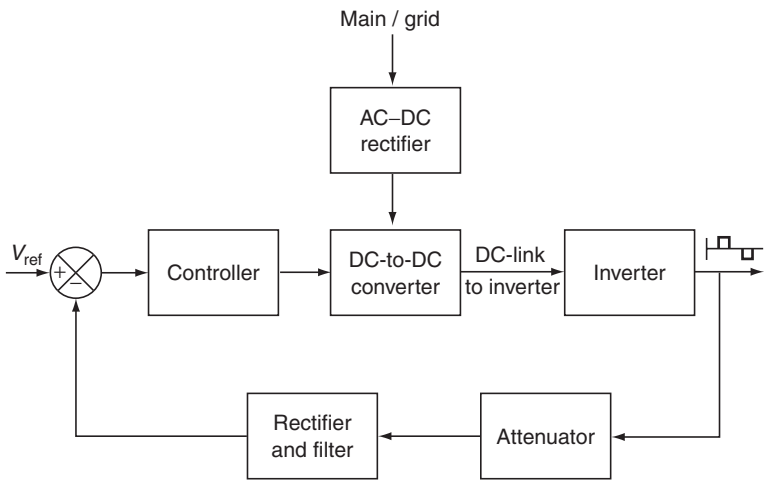


Figure 6.20 | Quasi-square wave regulation based on DC-link voltage control.

1. As the amplitude is a function of the DC link, the output can be controlled by controlling the DC-link voltage.
2. As the amplitude is also a function of the angle θ, the output can be controlled by controlling θ, which is also called variable duty cycle control.

The block schematics for both these types of control are shown in Figures 6.20 and 6.21, respectively. Figure 6.20 shows the block schematic of a DC-link control for a quasi-square wave inverter. Here the average of the output is controlled by controlling the DC-link voltage. The output voltage is rectified and filtered to obtain the average value of the output voltage waveshape. This voltage is compared with a desired reference value for the output average, V_{ref}. The error is passed through a controller like a proportional or a proportional–integral controller which is used to modulate the duty cycle of the switches in the DC–DC converter. The DC–DC converter gets its energy input from the output of a grid/mains rectified capacitor-filter circuit. The output voltage of the DC–DC converter is varied through duty cycle control which in turn feeds the quasi-square wave inverter, thereby controlling the average value of the output quasi-square wave.

In Figure 6.21 the block diagram for duty cycle control is shown, wherein the angle α is varied. The control of this angle is obtained by comparing a DC signal, which in this case is the controller output with

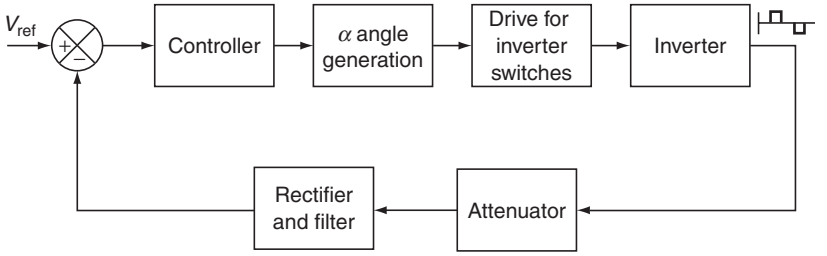


Figure 6.21 | Quasi-square wave regulation based on duty cycle control.

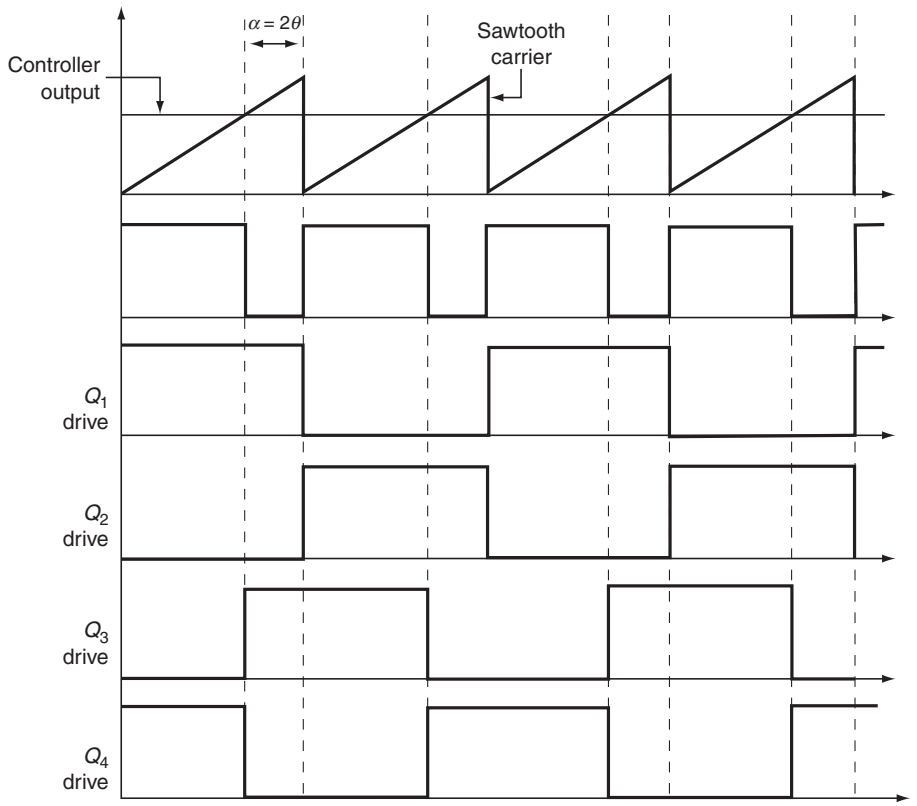


Figure 6.22 | Drive signal generation by duty cycle control for a quasi-square wave full-bridge inverter.

a ramp as shown in the Figure 6.22. The α angle generation block converts the controller output to duty cycle variations and generates the drive signals for the switches of the full-bridge inverter.

In the case of the DC-link control, the power is handled by both the DC–DC converter and the inverter. Therefore, this method needs two power blocks of the same rating. Further, this also implies that there are losses in the DC–DC converter and also the inverter, thereby giving reduced efficiency as compared to the variable duty ratio control wherein there is only one power block. Thus, the variable duty cycle control is more common in practice.

6.5 Three-Phase Inverter Topologies

The three-phase inverter topologies require three-arm bridge structures. They can be built using three two-level single-arm bridge topologies as shown in Figure 6.2(b). Consider three two-level single-arm bridge configuration inverters joined together as shown in Figure 6.23. Merging the input supplies of the three single-phase bridges together, one has the system as shown in Figure 6.24 wherein the load becomes a three-phase *star* topology. The junction of the three-phase *star* load is the neutral point N. The neutral (N) of the load is connected to the center of the split supply (O) which is called the *pole voltage*. Therefore, it may be said that the *load neutral is connected to the pole voltage*. In Figure 6.24, i_n is the neutral current. However, in a balanced

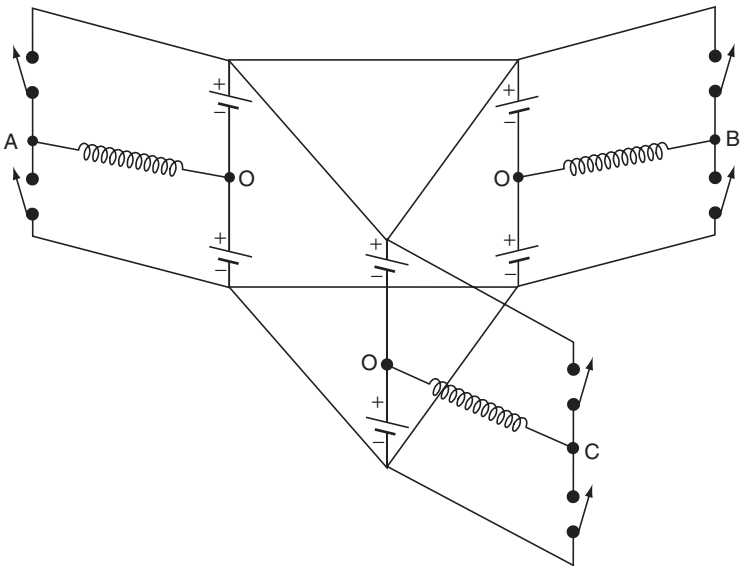


Figure 6.23 Three two-level single-arm bridge configurations used to form a three-phase inverter.

three-phase system, the neutral current will be zero and therefore reduces to the configuration as shown in Figure 6.25. In the configuration shown in Figure 6.25, the load neutral point is floating wherein the neutral current is forced to zero for a *star* load. The load may be a *star* or an equivalent *delta* load.

Referring to Figure 6.24, the potential at "N" with respect to "O" can be determined by knowing the potentials of A, B, and C with respect to O:

$$V_{AN} = V_{AO} - V_{NO} \tag{6.23}$$

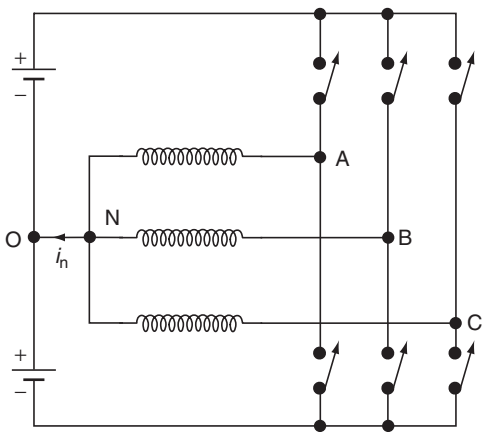


Figure 6.24 Three-phase inverter topology with neutral connected to pole (center of split supply).

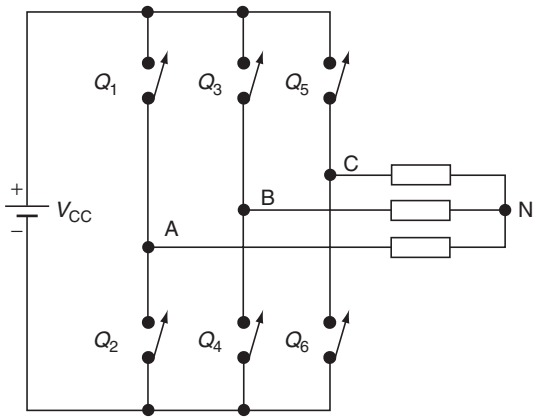


Figure 6.25 | Three-phase inverter topology with neutral floating.

$$V_{BN} = V_{BO} - V_{NO} \tag{6.24}$$

$$V_{CN} = V_{CO} - V_{NO} \tag{6.25}$$

Adding Eqs. (6.23)–(6.25), one obtains

$$V_{AN} + V_{BN} + V_{CN} = V_{AO} + V_{BO} + V_{CO} - 3V_{NO} \tag{6.26}$$

But $V_{AN} + V_{BN} + V_{CN} = 0$ at every instant. Hence from Eq. (6.26),

$$V_{NO} = \frac{V_{AO} + V_{BO} + V_{CO}}{3} \tag{6.27}$$

Knowing the phase to center-point voltages, V_{AO}, V_{BO} and V_{CO}, the neutral-to-center point voltage V_{NO} can be obtained from Eq. (6.27). The phase voltages V_{AN}, V_{BN} and V_{CN} can then be estimated using Eqs. (6.23)–(6.25). The above analysis is valid even for the configuration of Figure 6.25. Here the center point O is assumed as an imaginary or virtual point with a potential of half the input DC voltage.

Six-Step Inverter

The six-step inverter is a three-phase inverter and as the name suggests, there are six steps in the phase-voltage waveforms per cycle. The phase-voltage waveform for phase AN is shown in Figure 6.26. The drive signals for the various switches of the three-phase bridge inverter topology are also depicted in Figure 6.26.

Based on the switching sequence for each arm of the inverter, the equivalent circuits for the load at various switching instants are depicted in Figure 6.27. The entire cycle is divided into six operative modes. Each mode is operative for 60° of the whole cycle. During mode 1, Q_1, Q_4 and Q_5 are ON and the other three switches are OFF. The equivalent circuit for the load phases is as shown in Figure 6.27 for mode 1. The equivalent circuits given in Figure 6.27 at various intervals of times indicate the manner in which the phase-voltage waveform changes at various instants of time. From Figure 6.27, it is seen that in mode 1, both phases A and C are connected to V_{CC} and the phase B is connected to ground point g_{nd}. Therefore, two-third of V_{CC} drops across phase B and one-third of V_{CC} drops across phases A and C with respect to the neutral

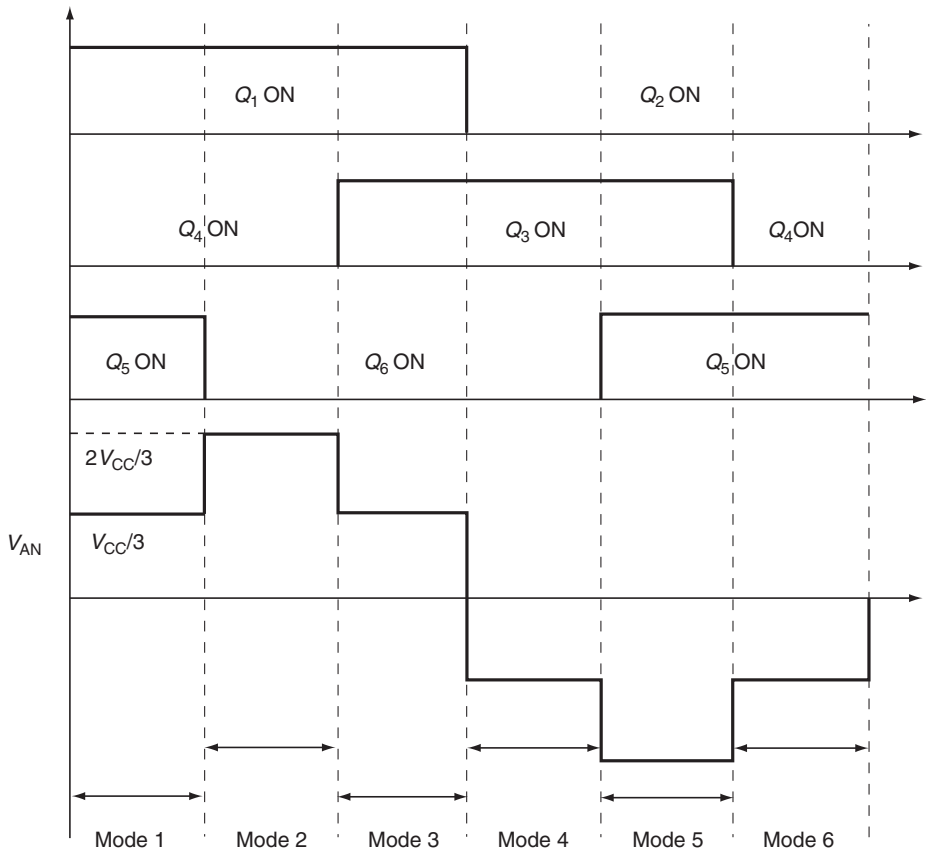


Figure 6.26 Switching sequence for the 3ϕ inverter switches to generate the six-step phase waveform.

point N as phases A and C are in parallel. By similar argument, the voltage values for all the other modes can be obtained from Figure 6.27.

It can be observed that each arm is a two-level arm that produces only square waveforms. However, the switching sequence for the switches in each arm is such that the square waveforms of the arms are displaced with respect to each other by 120°. *Thus it can be argued that if each arm generates a square wave and is displaced with the waveform of the other two arms by 120° as depicted in Figure 6.27, then a six-step waveform is obtained across the phases of the three-phase load.*

Current-Controlled Inverter

Another type of three-phase inverter in common use is the current-controlled pulse-width modulated (PWM) inverter that consists of a PWM voltage source inverter fitted with current-regulating loops to provide a controlled current output (pulse-width modulation is discussed in greater detail in the next section). It is generally used in AC motor applications. Normally, a sinusoidal reference current waveform, i_{ref}, is generated and fed to a comparator, together with the actual measured current of the motor. The simple approach

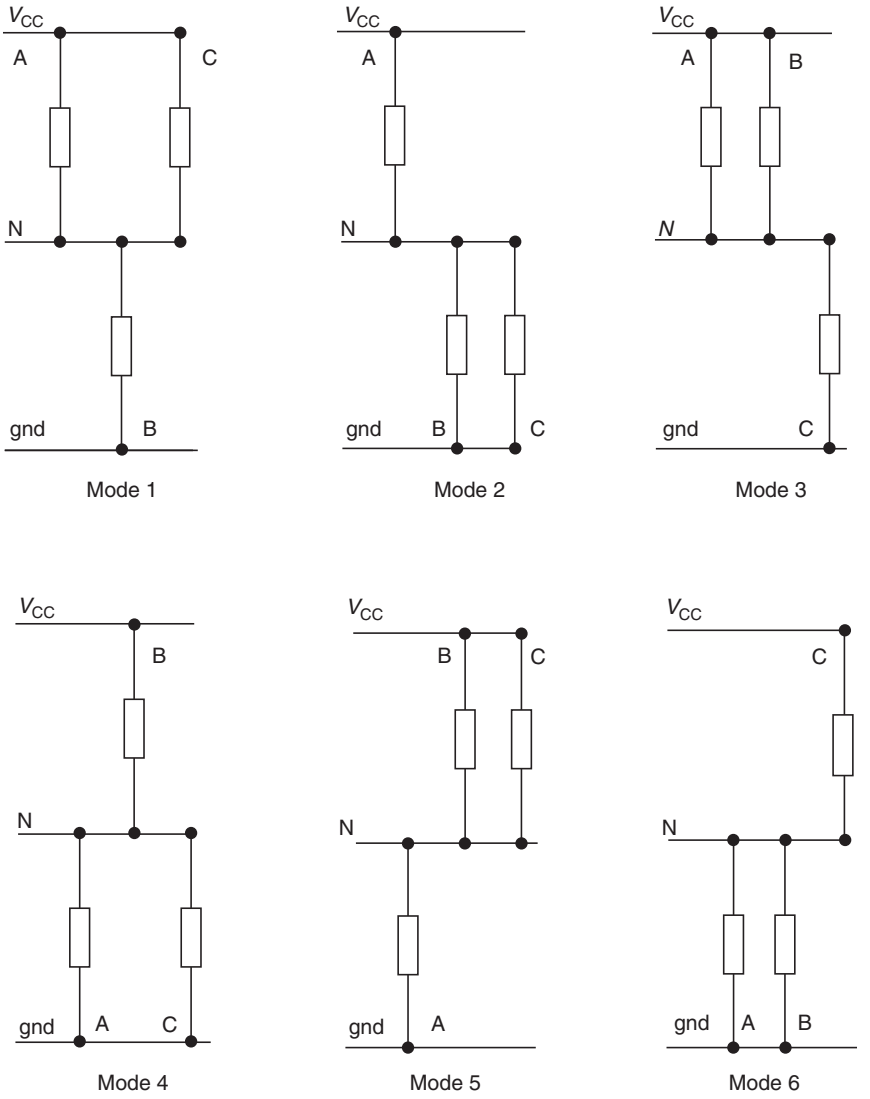


Figure 6.27 | Equivalent circuits for the various modes of the six-step inverter.

uses the comparator error to switch the devices in the inverter, so as to limit the instantaneous current error. Figure 6.28 shows the control for one inverter leg. If the motor phase current is more positive than the reference current value, the upper device is turned OFF and the lower device is turned ON, causing the motor current to decrease and vice-versa. The comparator is a hysteresis comparator that determines the permitted deviation of the actual phase current from the reference value before an inverter switching is initiated. Thus, the actual current tracks the reference current without significant amplitude error or phase delay. In three-phase systems, there is usually an independent current controller for each inverter phase.

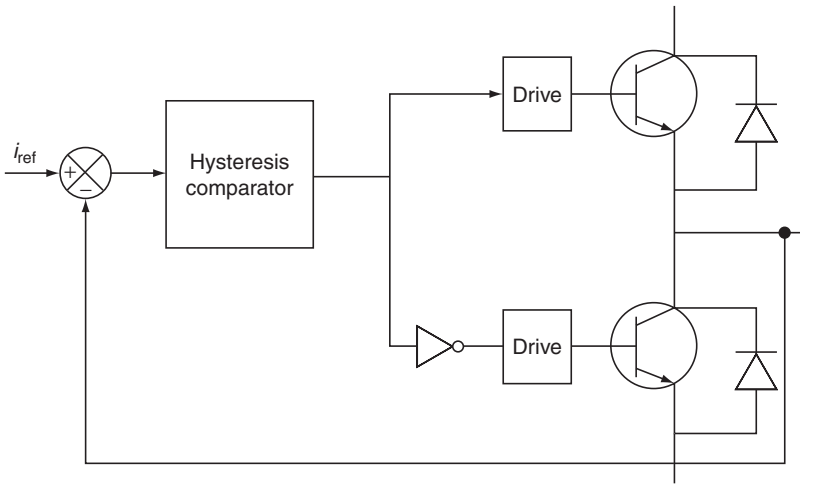


Figure 6.28 A current control scheme showing one leg of a three-phase inverter.

Figure 6.29 illustrates the type of output current waveform obtained with the simple hysteresis or ON/OFF current controller. A small dead band gives a near sinusoidal motor current with a small current ripple, but results in a high switching frequency in the inverter. One should note that the switching frequency is not constant for a given dead band but is modulated by the variations in motor inductance and back emf. When the back emf of the motor is low, the switching frequency may rise excessively. In a three-phase system without a neutral connection, the instantaneous current error can reach double the hysteresis band. In addition, the variable switching frequency produces objectionable acoustic noise and despite its simplicity this ON/OFF or "bang–bang" technique is seldom used in practice. However, this principle is used in certain

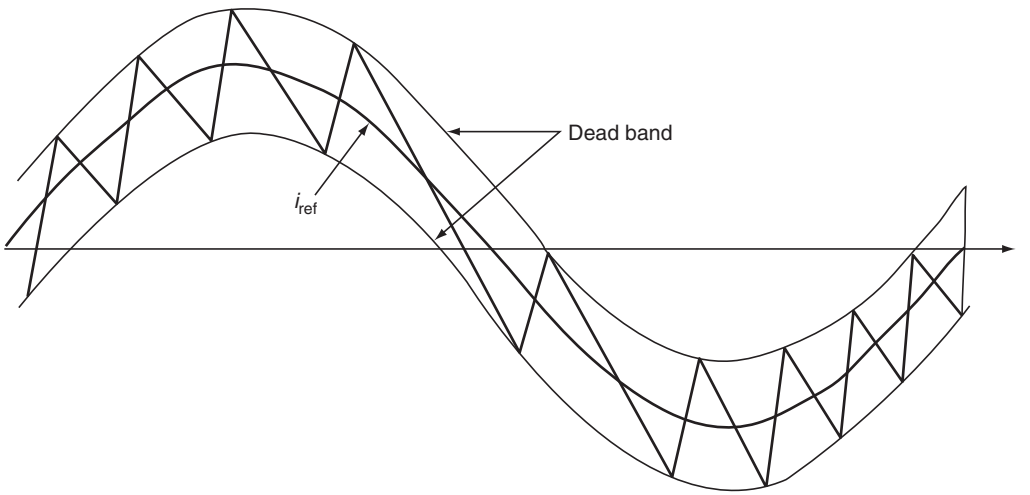


Figure 6.29 Typical phase-current waveshape of the current hysteresis controller.

AC motor control applications called *direct torque control*. Alternately the vector control technique is also used for current control applications. This is discussed in Chapter 10.

A fixed switching frequency is preferred because the acoustic noise is less and the inverter switching losses are more predictable. Figure 6.30 shows the current control technique in which the current error is passed through a sample-and-hold (S/H) circuit and compared with a fixed frequency triangular carrier wave. The current error is essentially the reference or the modulating signal. The resulting PWM signal, whose duty cycle is such that the current error tends towards zero, controls the inverter switching.

6.6 Pulse-Width Modulation

In the previous sections, the various topologies of the inverters starting from self-driven inverters to three-phase inverters were discussed. Some of the common topologies like the push–pull, half-bridge and full-bridge inverter structures were also discussed. Obtaining simple inverter output waveshapes like the quasi-square wave in the case of single-phase inverters and the six-step waveshape in the case of the three-phase inverters were discussed. However, for more complex output waveshapes, the inverter switches are controlled by means of PWM drive waveforms. The nature and the quality of the output waveshape is controlled by the switching sequence that is applied to the inverter switches. The main objective of the pulse-width modulation is to *control* the fundamental component of the PWM waveform with respect to the applied DC voltage. This control on the fundamental component is expressed in terms of a measure called the *modulation control index m*. Higher the value of "*m*" lower will be the relative harmonic content. As *m* decreases, there will be an increase in the relative harmonic content. The *modulation control index* is defined as

$$m = \frac{V_1}{V_{dc}} \tag{6.28a}$$

where V_1 is the peak value of the fundamental component of the PWM waveform; V_{dc} the value of the applied input DC to the inverter.

For the harmonic content, a measure called the harmonic index is defined for each of the harmonic component relative to the fundamental. The harmonic index h_n for the nth harmonic is defined as

$$h_n = \frac{V_n}{V_1} \tag{6.28b}$$

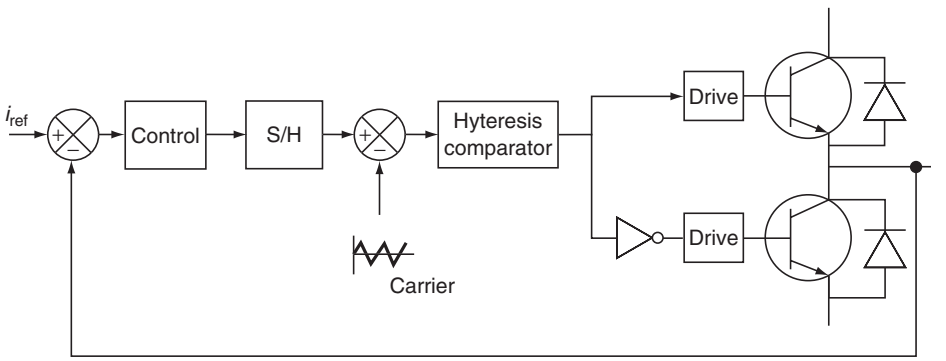


Figure 6.30 | Carrier-based current control.

where V_n is the peak value of the nth harmonic component of the PWM waveform; V_1 the peak value of the fundamental of the PWM waveform.

For a practical system, the voltage generated by the inverter should fulfill the following two constraints:

1. The negative part of the waveform should be a mirror reflection of the positive part of the waveform. This means that the generated waveform should be symmetrical about the time axis. This will ensure that the even harmonics are absent in the resulting PWM waveform.
2. In the case of three-phase systems, the waveforms generated should be symmetrical. This means that the waveform of each phase should have the same amplitude as the other phases but should be displaced by 120° with respect to the other two phases.

The PWM can be categorized into the following two general classes, namely,

1. PWM without harmonic elimination;
2. PWM with harmonic elimination.

In the former class, the PWM is performed without any formal effort to cancel specific harmonics whereas in the latter class, the PWM is performed such that specific harmonics can be eliminated. This section discusses the PWM without harmonic elimination and the following sections will discuss on the issue of PWM with specific harmonic elimination.

The PWM without harmonic elimination can be further classified into the following types:

1. quasi-square modulation;
2. end pulse modulation;
3. center pulse modulation;
4. sinusoidal pulse-width modulation;
5. PWM by phase modulation;
6. space-vector modulation.

The Fourier series representation of the waveforms will be used to analyze and discuss the various modulation methods. The Fourier series representation of any general waveform is given by

$$f(\omega t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + \sum_{n=1}^{\infty} b_n \sin(n\omega t) \tag{6.29}$$

where

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(\omega t) \mathrm{d}\omega t$$

$$a_n = \frac{2}{2\pi} \int_0^{2\pi} f(\omega t) \cos(n\omega t) \mathrm{d}\omega t$$

$$b_n = \frac{2}{2\pi} \int_0^{2\pi} f(\omega t) \sin(n\omega t) \mathrm{d}\omega t$$

Here a_0 gives the measure of the average value of the waveform. In the discussions to follow, as the waveforms that are considered are AC waveforms with zero average, a_0 value is considered to be zero. The waveforms that are considered in the sections to follow have quarter wave symmetry. This means that

the waveform is symmetrical about a vertical line at $\theta = \pi/2$. In such cases, a_n or b_n can be evaluated considering only quarter of the periodic waveform. Thus for waveforms with quarter wave symmetry,

$$a_n = \frac{2}{\pi/2}\int_0^{\pi/2} f(\omega t)\cos(n\omega t)\mathrm{d}\omega t$$

$$b_n = \frac{2}{\pi/2}\int_0^{\pi/2} f(\omega t)\sin(n\omega t)\mathrm{d}\omega t$$

The waveforms that are discussed in the sections to follow have the negative parts that are a mirror reflection of the positive parts. Therefore the even harmonics are absent. This implies that the harmonic number n will take only odd values like 1, 3, 5, 7, 9, etc.

The PWM patterns that will be discussed in the following sub-sections and the next section are generally obtained by programming the timers in a digital processor. Either a microcontroller or a digital signal processor is generally used for the purpose. However, the PWM sequence can also be obtained by comparison of a sawtooth or a triangular wave carrier with an appropriate analog signal to generate the various PWM patterns. The PWM signal is given as input to the drive circuitry of the power switches of the inverter. A block schematic of a typical PWM inverter is shown in Figure 6.31. In Figure 6.31, the output waveshape that is pulse-width modulated is filtered to obtain a sinusoidal output voltage. This voltage is fed back for control purposes. It is rectified and filtered to obtain the average value of the output waveform. This DC voltage is compared with a desired reference voltage. The error voltage is passed through a proportional–integral controller.

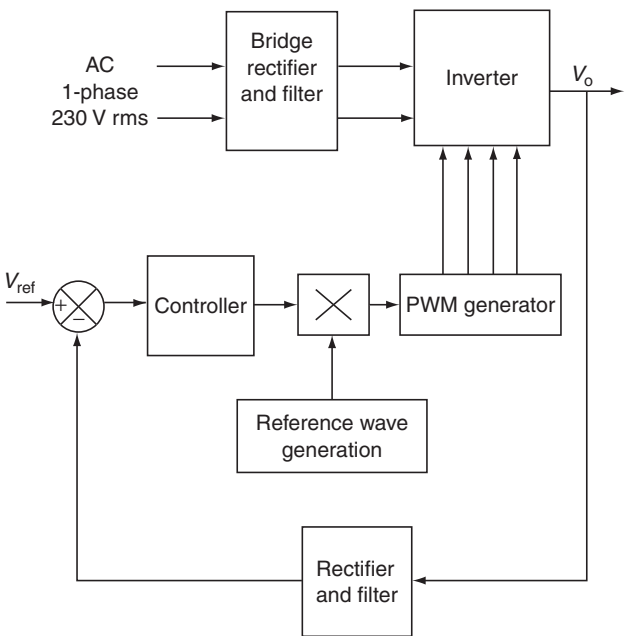


Figure 6.31 | Typical block schematic of a PWM inverter.

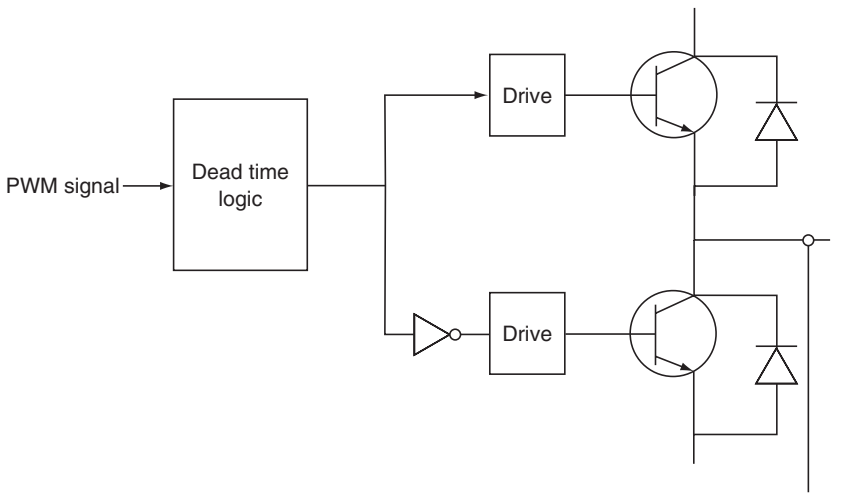


Figure 6.32 | PWM signal driving one arm of an inverter.

The controller output is used for controlling the amplitude of a reference wave. This controlled sine wave reference is compared with a carrier frequency waveform to generate the PWM waveform that is used for switching the inverter switches.

Figure 6.32 illustrates the schematic for driving one arm of the inverter bridge. The PWM signal that is generated by a specific algorithm is passed through a dead time logic. This logic ensures that at every voltage transition of the PWM pattern, sufficient dead time is provided for the ON-state switch to turn OFF and the OFF-state switch to subsequently turn ON, such that the positive rail of the inverter bridge is not short-circuited to the negative rail. The signal is then used to drive the inverter switches of a specific arm. For single-phase inverters, there are two bridge arms wherein one bridge arm gets a PWM signal that is generated by a reference modulating wave that is phase shifted by 180° with respect to the reference modulating wave for the other arm.

In the case of three-phase inverters, there are three bridge arms, one each for phases A, B and C. Each arm is given a PWM signal wherein the PWM signal is generated by a reference modulating wave that is phase shifted 120° with respect to the reference modulating waves of the other two arms.

Quasi-Square Modulation

Quasi-square modulation is the same as the duty cycle control as discussed in the section "Quasi-Square Wave Inverter". The quasi-square waveform is again depicted here in Figure 6.33 wherein V_{dc} is the applied DC voltage to the inverter and δ is the angular pulse width for the quasi-square wave. The pulse width δ is varied or modulated to obtain the variation in modulation index and the harmonic content.

The Fourier series representation of the quasi-square waveforms is given by Eq. (6.31). Referring to Figure 6.33, it can be observed that the waveform has odd symmetry about the vertical axis at 0. The waveform further has quarter wave symmetry about $\theta = \pi/2$; therefore

$$b_n = \frac{2}{\pi/2}\int_0^{\theta} 0 \sin(n\omega t)\mathrm{d}\omega t + \frac{2}{\pi/2}\int_{\theta}^{\pi/2} V_{dc} \sin(n\omega t)\mathrm{d}\omega t$$

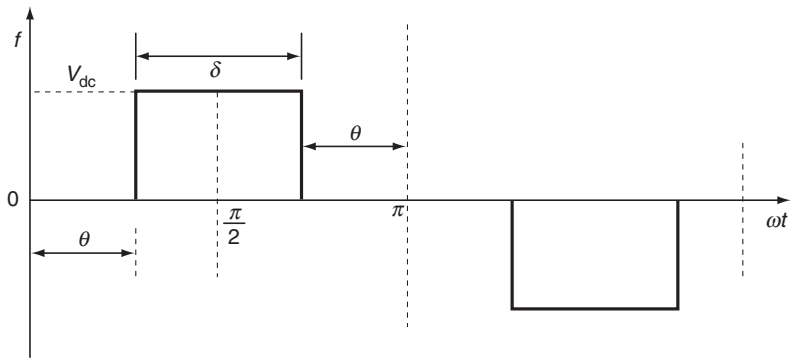


Figure 6.33 | Quasi-square wave.

$$b_n = \frac{4V_{dc}}{n\pi}[-\cos(n\omega t)|_{\theta}^{\pi/2}]$$

$$b_n = \frac{4V_{dc}}{n\pi}[\cos n\theta - \cos(n\pi/2)]$$

Now $\cos(n\pi/2) = 0$ for $n = 1, 3, 5, 7, \ldots$. The waveform does not contain even harmonics as the positive and negative parts are mirror images. Thus one obtains

$$b_n = \frac{4V_{dc}}{n\pi}\cos n\theta \text{ for } n = 1, 3, 5, \ldots \tag{6.30}$$

$$f_{quasi} = \frac{4V_{dc}}{n\pi}\sum_{n=1}^{\infty}[\cos(n\theta)\sin(n\omega t)] \text{ for } n = 1, 3, 5, \ldots \tag{6.31a}$$

The angle,

$$\theta = \frac{\pi}{2} - \frac{\delta}{2} \tag{6.31b}$$

It can be observed from Eq. (6.31a) that for $\theta = 0$, f_{quasi} becomes a square waveform f_{square} that is given as

$$f_{square} = \frac{4V_{dc}}{n\pi}\sum_{n=1}^{\infty}[\sin(n\omega t)] \text{ for } n = 1, 3, 5, \ldots$$

The peak value of the fundamental component of f_{quasi} is given as

$$V_1 = \frac{4V_{dc}}{\pi}\cos(\theta) = \frac{4V_{dc}}{\pi}\cos\left(\frac{\pi}{2} - \frac{\delta}{2}\right) = \frac{4V_{dc}}{\pi}\sin\left(\frac{\delta}{2}\right) \tag{6.32}$$

From Eq. (6.32), the modulation control index m is obtained as

$$m = \frac{V_1}{V_{dc}} = \frac{4}{\pi}\sin\left(\frac{\delta}{2}\right) \tag{6.33}$$

The harmonic indices are given as

$$h_n = \frac{V_n}{V_1} = \frac{\left|\cos(n\theta)\right|}{n\sin(\delta/2)} \text{ for } n = 3, 5, 7, \ldots \tag{6.34}$$

The modulation control index and the harmonic indices are dependent on the angular pulse width δ and therefore θ. The plot of the modulation control index and the harmonic indices are shown in Figure 6.34. The third harmonic and the harmonics that are multiples of three are not plotted as they will be eliminated in a three-phase system. It can be observed that at low modulation control index (m) the relative harmonic content is large. For values of m less than 0.7, the relative harmonic content starts becoming significantly high. As δ tends towards π, the modulation control index tends towards that of a square wave.

For three-phase circuits, the phase-to-neutral voltages are quasi-square waveforms. Each phase waveform is displaced by 120° with respect to the other two phase waveforms. The line-to-line voltage waveform is the difference between two phase-voltage waveforms and is as depicted in Figure 6.35. Figure 6.35 shows the A phase and B phase voltage. The line-to-line voltage between A and B phases is also shown. The line-to-line voltage waveforms contain the same harmonics as the phase-voltage waveforms; however, the third harmonic and the harmonics that are multiples of three are eliminated. Further, the line-to-line voltage waveforms are $\sqrt{3}$ times that of the phase-to-neutral waveforms. Thus

$$f_{\text{quasi-}3\phi} = \frac{4\sqrt{3}V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}\cos(n\theta)\sin(n\omega t) \text{ for } n = 1, 5, 7, 11, \ldots$$

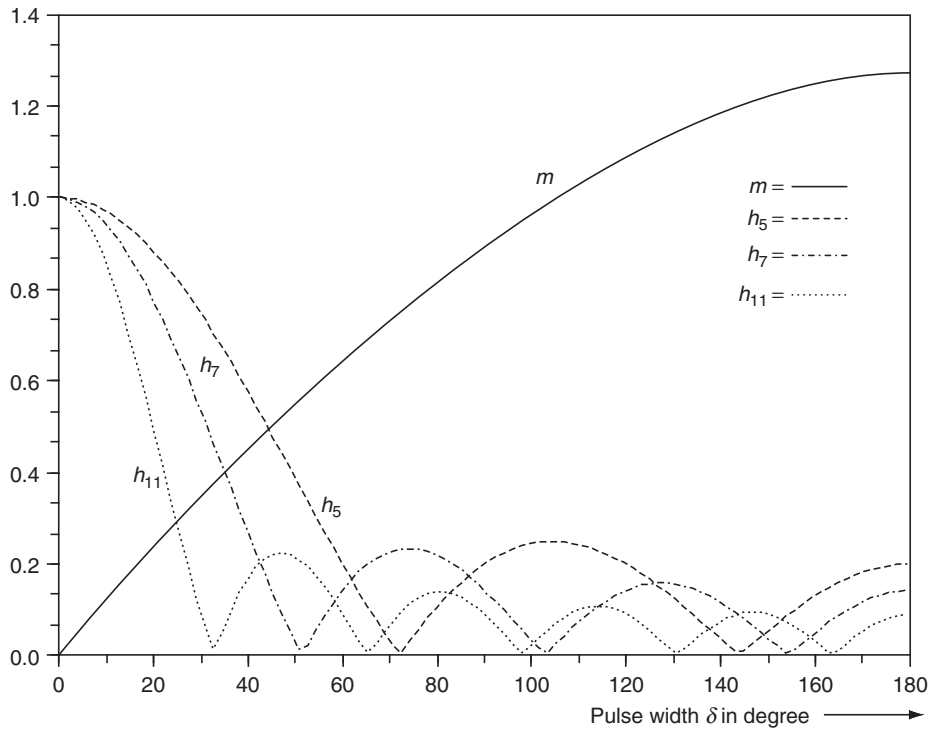


Figure 6.34 | Modulation control index and harmonic indices for quasi-square modulation.

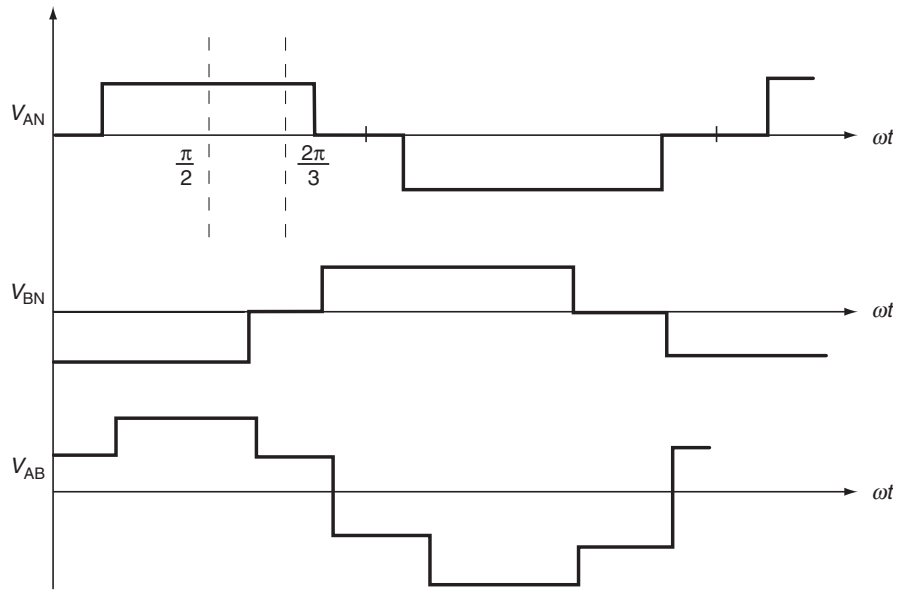


Figure 6.35 | Phase- and line-voltage waveforms for quasi-square modulation.

End Pulse Modulation

End pulse modulation is another pattern of modulation wherein each half-cycle starts and ends with an interval where the voltage is of opposite sign. The end pulse waveshape is shown in Figure 6.36. The end pulse width θ is varied or modulated to obtain the variation in modulation control index and the relative harmonic content.

The Fourier series representation of the end pulse waveform is given by Eq. (6.36). Referring to Figure 6.36, it can be observed that the waveform has odd symmetry about the vertical axis at 0. The waveform further has quarter wave symmetry about $\theta = \pi/2$; therefore only b_n coefficients exist. Thus,

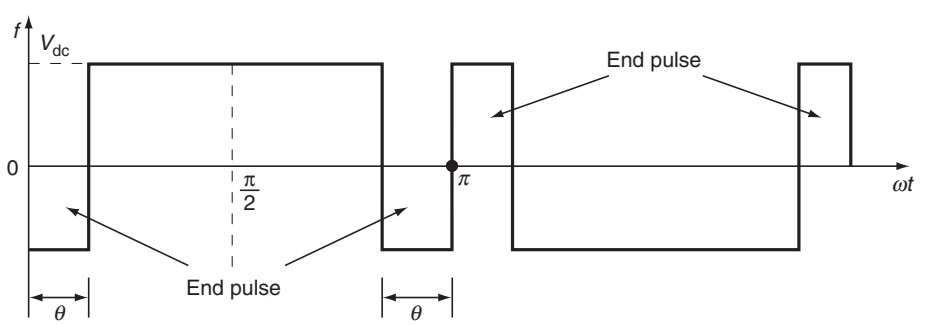


Figure 6.36 | End pulse waveform.

$$b_n = \frac{2}{\pi/2}\int_0^{\theta}(-V_{\text{dc}})\sin(n\omega t)\text{d}\omega t + \frac{2}{\pi/2}\int_{\theta}^{\pi/2} V_{\text{dc}}\sin(n\omega t)\text{d}\omega t$$

$$b_n = \frac{-4V_{\text{dc}}}{n\pi}\left[-\cos(n\omega t)\Big|_0^{\theta}\right] + \frac{4V_{\text{dc}}}{n\pi}\left[-\cos(n\omega t)\Big|_{\theta}^{\pi/2}\right]$$

$$b_n = \frac{4V_{\text{dc}}}{n\pi}(2\cos n\theta - 1) \text{ for } n = 1, 3, 5, 7, \ldots \tag{6.35}$$

$$f_{\text{endpulse}} = \frac{4V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}(2\cos n\theta - 1)\sin(n\omega t) \text{ for } n = 1, 3, 5, \ldots \tag{6.36}$$

It can be observed from Eq. (6.36) that for $\theta = 0$, f_{endpulse} becomes a square waveform f_{square} that is given as

$$f_{\text{square}} = \frac{4V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}\sin(n\omega t) \text{ for } n = 1, 3, 5, \ldots$$

The peak value of the fundamental component of f_{endpulse} is given as

$$V_1 = \frac{4V_{\text{dc}}}{\pi}(2\cos\theta - 1) \tag{6.37}$$

From Eq. (6.37), the modulation control index (m) is obtained as

$$m = \frac{V_1}{V_{\text{dc}}} = \frac{4}{\pi}(2\cos\theta - 1) \tag{6.38}$$

The harmonic indices are given as

$$h_n = \frac{V_n}{V_1} = \frac{\left|2\cos n\theta - 1\right|}{n(2\cos\theta - 1)} \tag{6.39}$$

where $n = 3, 5, 7, \ldots$. The plot of the modulation control index and the harmonic indices are shown in Figure 6.37. The relative harmonics that are not multiples of three are plotted. It can be observed that as the modulation control index (m) decreases, the relative harmonic content increases. For values of m less than 1.1, the relative harmonic content starts becoming significantly higher. As θ tends towards $\pi/6$ the modulation control index drops to zero as can be seen from Eq. (6.38). It can be observed by comparing Figures 6.34 and 6.37 that the range of the modulation control index is larger for the quasi-square modulation for a given relative harmonic content.

For three-phase circuits, the phase-to-neutral voltages are the end pulse waveforms as depicted in Figure 6.38. Each phase waveform is displaced by 120° with respect to the other two-phase waveforms. The line-to-line voltage waveform is the difference between two-phase voltage waveforms and is as depicted in Figure 6.38. Figure 6.38 shows the A phase and B phase voltage. The line-to-line voltage between A and B phases is also shown. The line-to-line voltage waveforms contain the same harmonics as the phase-voltage waveforms but the third harmonic and the harmonics that are multiples of three are eliminated. The Fourier series expansion for the line-to-line voltage waveforms is given as

$$f_{\text{endpulse-3}\phi} = \frac{4\sqrt{3}V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}(2\cos n\theta - 1)\sin(n\omega t) \text{ for } n = 1, 5, 7, 11, \ldots$$

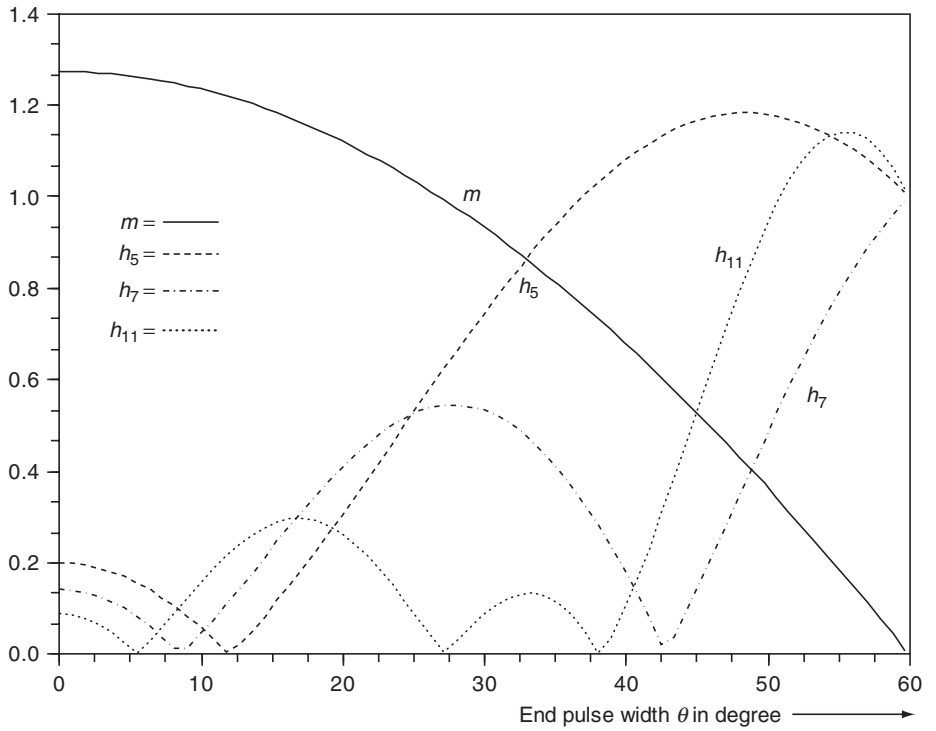


Figure 6.37 | Modulation control index and harmonic indices for end pulse modulation.

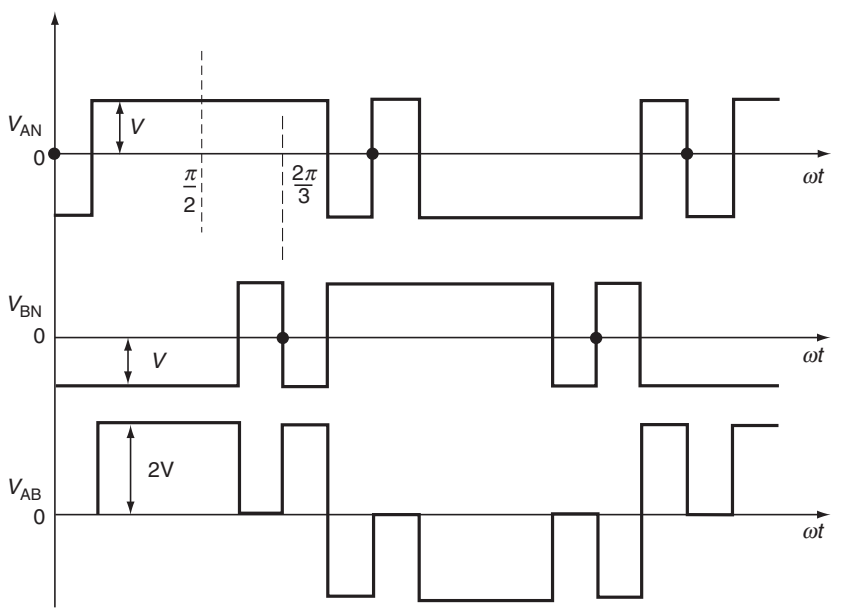


Figure 6.38 | Phase- and line-voltage waveforms for end pulse modulation.

Center Pulse Modulation

Center pulse modulation is another variant of modulating the square wave wherein each half-cycle consists of an interval in the center of the half-cycle where the voltage is of opposite sign. The center pulse waveshape is shown is Figure 6.39. The center pulse width δ is varied or modulated to obtain the variation in modulation control index and the relative harmonic content.

The Fourier series representation of the center pulse waveform is given by Eq. (6.41). Referring to Figure 6.39, it can be observed that the waveform has odd symmetry about the vertical axis at 0. Therefore the Fourier coefficients $a_n = 0$. The waveform further has quarter wave symmetry about $\theta = \pi/2$, therefore

$$b_n = \frac{2}{\pi/2}\int_0^{\theta} V_{\text{dc}} \sin(n\omega t)\,\mathrm{d}\omega t + \frac{2}{\pi/2}\int_{\theta}^{\pi/2} (-V_{\text{dc}})\sin(n\omega t)\,\mathrm{d}\omega t$$

$$b_n = \frac{-4V_{\text{dc}}}{n\pi}\left[-\cos(n\omega t)\Big|_0^{\theta}\right] + \frac{4V_{\text{dc}}}{n\pi}\left[-\cos(n\omega t)\Big|_{\theta}^{\pi/2}\right]$$

$$b_n = \frac{4V_{\text{dc}}}{n\pi}(1 - 2\cos n\theta) \text{ for } n = 1, 3, 5, 7, \ldots \tag{6.40}$$

$$f_{\text{centerpulse}} = \frac{4V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}(1 - 2\cos n\theta)\sin(n\omega t) \text{ for } n = 1, 3, 5, \ldots \tag{6.41a}$$

$$\delta = \pi - 2\theta \tag{6.41b}$$

It can be observed from Eq. (6.41a) that for $\theta = \pi/2$, $\delta = 0$ and $f_{\text{centerpulse}}$ becomes a square waveform f_{square} that is given as

$$f_{\text{square}} = \frac{4V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}\sin(n\omega t) \text{ for } n = 1, 3, 5, \ldots$$

The peak value of the fundamental component of $f_{\text{centerpulse}}$ is given as

$$V_1 = \frac{4V_{\text{dc}}}{\pi}(1 - 2\cos\theta) = \frac{4V_{\text{dc}}}{\pi}\left[1 - 2\cos\left(\frac{\pi}{2} - \frac{\delta}{2}\right)\right] = \frac{4V_{\text{dc}}}{\pi}\left(1 - 2\sin\frac{\delta}{2}\right) \tag{6.42}$$

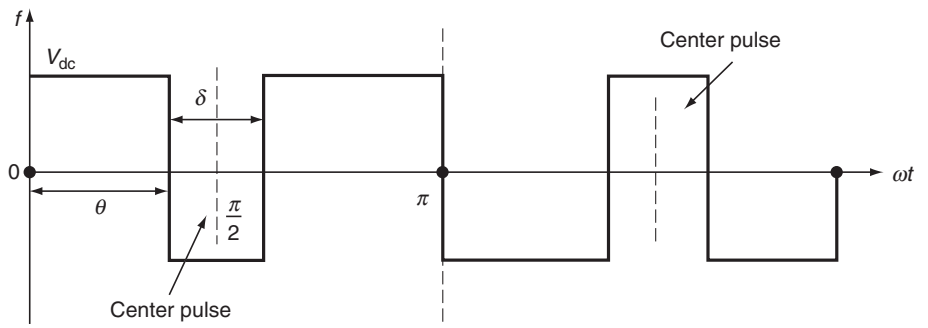


Figure 6.39 | Center pulse waveform.

From Eq. (6.42), the modulation control index m is obtained as

$$m = \frac{V_1}{V_{dc}} = \frac{4}{\pi}\left(1 - 2\sin\frac{\delta}{2}\right) \tag{6.43}$$

The harmonic indices are given as

$$h_n = \frac{V_n}{V_1} = \frac{\left|1 - 2\cos n\theta\right|}{n[1 - 2\sin(\delta/2)]} \tag{6.44}$$

where $n = 3, 5, 7, \ldots$. The plot of the modulation control index and the harmonic indices are shown in Figure 6.40. It can be observed that as the modulation control index m decreases, the relative harmonic content increases. For values of m less than 1.1, the relative harmonic content starts becoming significantly higher. The plot shows that there is a near linear relation between the modulation control index m and the pulse width δ. This implies that the control voltage and the fundamental of the output voltage are also linear.

The end pulse and center pulse modulation can have multiple pulses to cancel some specific harmonics. This would mean that there can be more than one end/center pulses. The harmonic elimination method uses multiple pulses to eliminate certain harmonics. This is discussed in the next section while discussing the pulse-width modulation with harmonic elimination.

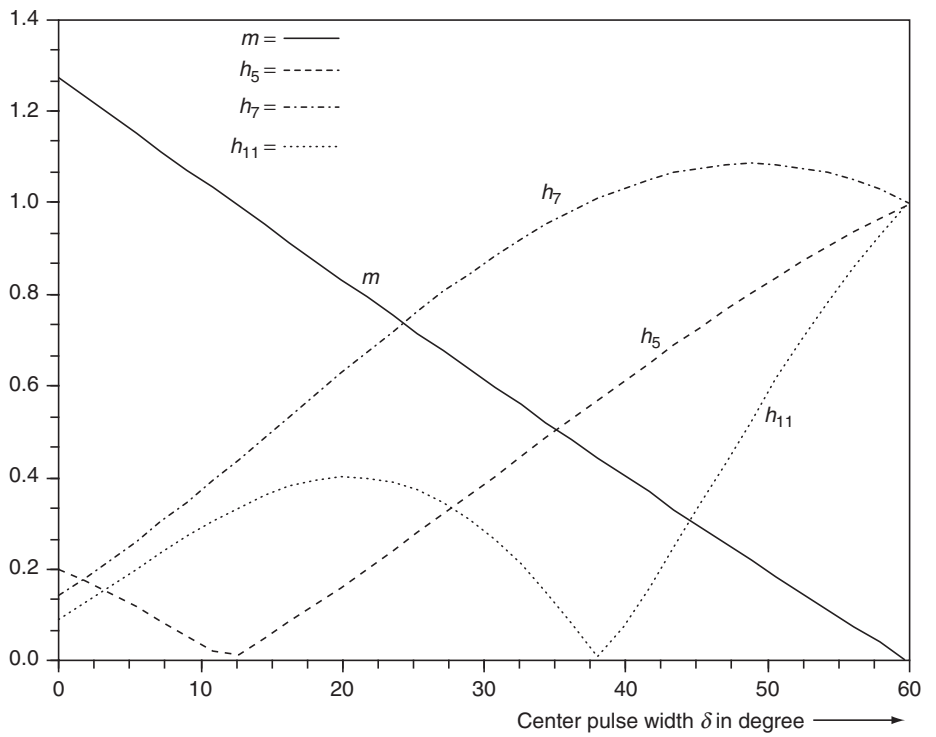


Figure 6.40 | Modulation control index and harmonic indices for center pulse modulation.

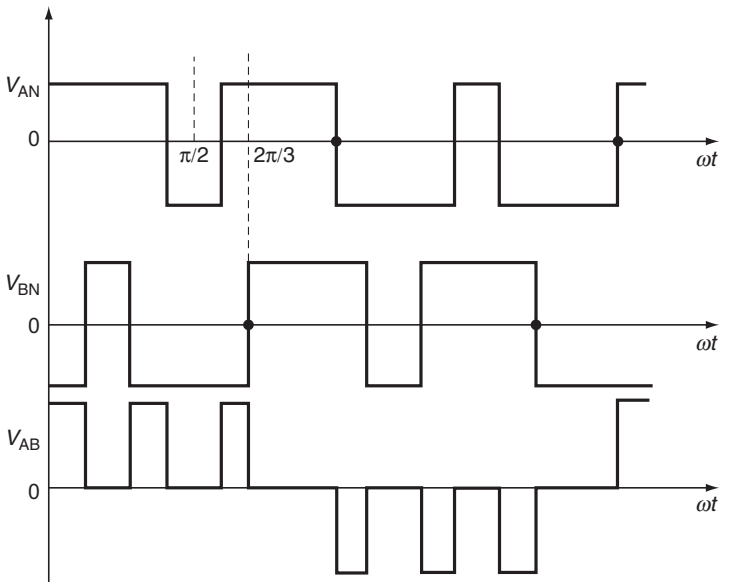


Figure 6.41 | Phase- and line-voltage waveforms for center pulse modulation.

For three-phase circuits, the phase-to-neutral voltages are the end pulse waveforms as depicted in Figure 6.41. Each phase waveform is displaced by 120° with respect to the other two-phase waveforms. The line-to-line voltage waveform is the difference between two-phase voltage waveforms and is as depicted in Figure 6.41. Figure 6.41 shows the phase A and phase B voltages. The line-to-line voltage between A and B phases is also shown. The line-to-line voltage waveforms contain the same harmonics as the phase-voltage waveforms but the third harmonic and the harmonics that are multiples of three are eliminated. The line-to-line voltage is expressed as

$$f_{\text{centerpulse-3}\phi} = \frac{4\sqrt{3}V_{\text{dc}}}{n\pi}\sum_{n=1}^{\infty}(1-2\cos n\theta)\sin(n\omega t) \text{ for } n = 1, 5, 7, 11, \ldots$$

Sinusoidal Pulse-Width Modulation

The sinusoidal output waveshape is one of the most common requirements for the loads. This is usually accomplished by means of driven inverter topologies. To obtain a sinusoidal output waveshape, the square, quasi, end or center PWM waveform that is generated from the inverter is passed through a filter such that the harmonics are attenuated and only the fundamental is allowed to pass through to the load. In the case of the square, quasi-square, end pulse or center pulse waveforms, the significant harmonic after the fundamental is the third harmonic in the case of single-phase loads and the fifth harmonic in the case of three-phase loads as discussed in the previous sub-sections. This implies that the filter should have a sharp cut-off profile to significantly attenuate the amplitudes of the third and higher harmonics.

In the sinusoidal pulse-width modulation technique, the pulse width of a high-frequency switching signal, called the carrier, is varied in accordance to the amplitude of a low-frequency signal, called the modulating signal, as shown in Figure 6.42.

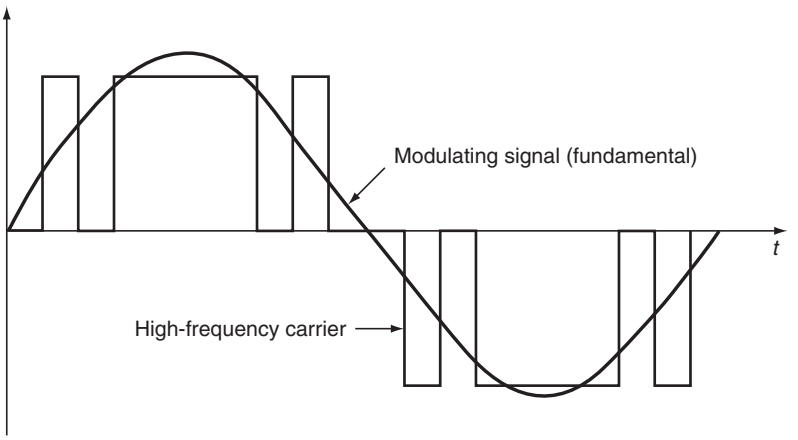


Figure 6.42 Unipolar pulse-width modulation: Sinusoidal fundamental modulating the high-frequency carrier.

It is seen from the frequency spectrum that the major harmonic after the modulating signal frequency is at the carrier frequency. If the modulating signal is a sinusoid of 50 Hz, then one gets a sinusoidal PWM waveform with a fundamental at 50 Hz. If the carrier frequency is 20 kHz, then the major harmonic after 50 Hz will be at 20 kHz. Therefore, to filter out the carrier from the PWM, the filter requirements are less stringent compared to filtering out the harmonics in a square or a quasi-square waveform. As the filter is supposed to filter out only the high-frequency carrier, the filter size will be small.

The sinusoidal PWM waveform as depicted in Figures 6.42 and 6.43 has quarter wave symmetry about the $\pi/2$ vertical and odd symmetry about the 0 vertical. Therefore, the Fourier coefficient $a_n = 0$ and b_n contains only odd harmonics.

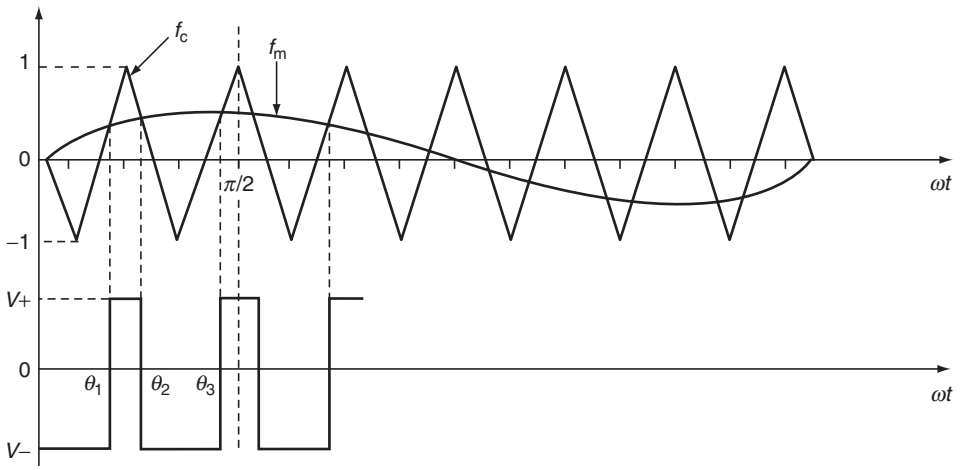


Figure 6.43 A portion of the bipolar sinusoidal pulse-width modulated waveform.

The angular pulse width is proportional to the amplitude of the modulating signal value at the specific θ values. The sinusoidal modulating signal is compared with a high-frequency triangle or sawtooth waveform. The points of intersection between the sinusoidal modulating signal and the triangle carrier are the angles at which the voltage transitions take place. Every voltage transition is called a "notch". The corresponding angles at which these transitions occur are called the "notch angles". The voltage transits between two levels from *V*+ to *V*– and vice-versa. *V*+ is usually the input DC voltage level V_{dc} and *V*– is usually $-V_{dc}$ or 0. At every notch, if the voltages transit between positive and negative voltage, then such a PWM is called bi-polar PWM as depicted in Figure 6.43. At every notch if the voltages transit between zero and either positive or negative voltage, then such a PWM is called a unipolar PWM as illustrated in Figure 6.42.

The carrier frequency or the frequency of the switching pulses is

$$\omega_c = 2\pi f_c; \quad T_c = \frac{1}{f_c}$$

and the modulating frequency or the frequency of the fundamental sinusoid is

$$\omega_m = 2\pi f_m; \quad T_m = \frac{1}{f_m}$$

The ratio of the carrier frequency to the modulating frequency is called the frequency ratio *r* that is given as

$$r = \frac{\omega_c}{\omega_m} = \frac{f_c}{f_m} \tag{6.45}$$

Referring to Figure 6.43, the Fourier representation of the sinusoidal PWM for "*p*" notches in a quarter cycle is given as

$$f_{\text{sinusoidal}} = \sum_{n=1}^{\infty} b_n \sin(n\omega t) \tag{6.46}$$

where

$$b_n = \frac{2}{\pi/2}\int_0^{\theta_1} V_+ \sin(n\omega t)\mathrm{d}\omega t + \frac{2}{\pi/2}\int_{\theta_1}^{\theta_2} V_- \sin(n\omega t)\mathrm{d}\omega t + \cdots + \frac{2}{\pi/2}\int_{\theta_p}^{\pi/2} V_+ \sin(n\omega t)\mathrm{d}\omega t$$

$$b_n = \frac{4V_+}{n\pi}\sum_{i=0}^{p} \cos n\theta_i - \cos n\theta_{i+1} + \frac{4V_-}{n\pi}\sum_{j=1}^{p} \cos n\theta_j - \cos n\theta_{j+1} \tag{6.47}$$

for *i* = even (0, 2, 4, 6, 8, …); *j* = odd (1, 3, 5, 7, …); *n* = odd (1, 3, 5, 7, …). Referring to Figure 6.43, the notch angles are determined by two waveforms: (a) the triangular carrier and (b) the modulating signal. The points of intersection of these two waveforms give the notch angles. Therefore, based on the modulation signal, the amplitude of the modulating signal at the notch angle is

$$f = M \sin\theta_i \tag{6.48}$$

where *M* is the *modulation index* which is defined as the ratio of the peak amplitude of the fundamental signal to the peak amplitude of the carrier signal. Based on the carrier signal and referring to Figure 6.43, the amplitude of the carrier signal at the notch angle is

$$f = (-1)^i \left(2i - \frac{2r}{\pi}\theta_i\right) \tag{6.49}$$

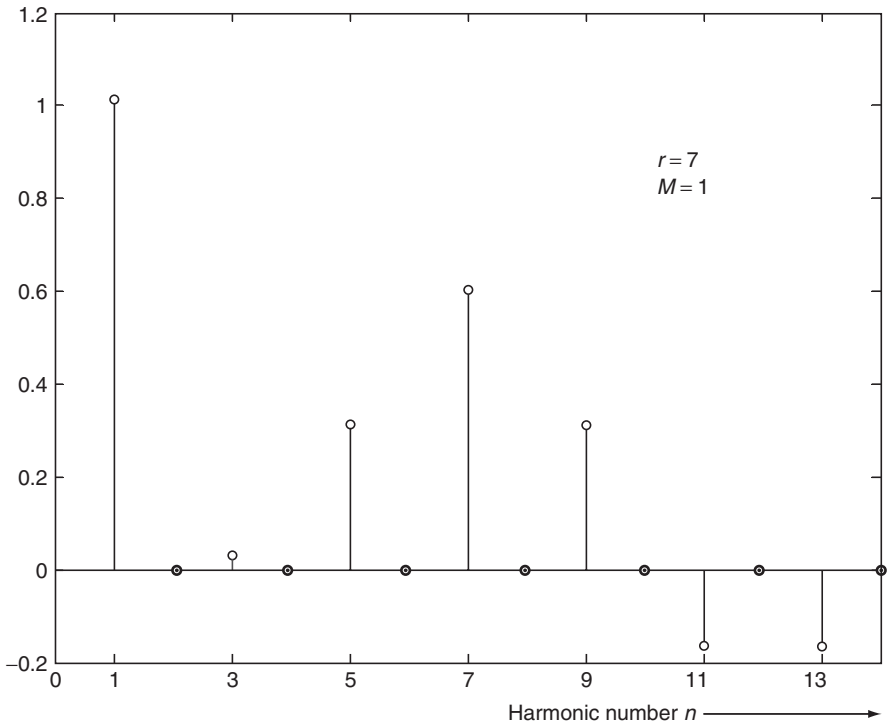


Figure 6.44 | Harmonic line spectrum for bipolar PWM.

where $2r/\pi$ is the slope of the triangular carrier. At the intersection points, Eqs. (6.48) and (6.49) will be equal. These two equations should be used to compute θ_i iteratively for the purpose of determining the notch angles for analysis. The notch angles thus determined can be substituted in Eq. (6.47) to obtain the various harmonic components. The harmonic line spectrum are plotted for a frequency ratio $r = 7$ and modulation index $M = 1$. Figure 6.44 shows the harmonic line spectrum for the bipolar PWM and Figure 6.45 shows the harmonic line spectrum for the unipolar PWM. It can be observed that the unipolar PWM gives a better modulation control index resulting in higher fundamental amplitude. The carrier amplitude is considerably reduced. Figure 6.46 shows the modulation control index m and the harmonic indices for the 5th, 7th and 11th harmonics for a bipolar PWM. Likewise, Figure 6.47 shows the modulation control index and the harmonic indices for the unipolar PWM. It can be observed that the bipolar PWM gives a better harmonic reduction than the unipolar PWM.

General Notes on Sinusoidal PWM

1. It should be observed that for low values of r, the relative harmonic content is significantly high. Care should be taken not to add even harmonics due to unsymmetry in the positive and negative half of the PWM pattern. The value of frequency ratio r should be an integer. This means that the carrier and the modulating waveforms are synchronous thereby avoiding discontinuities and harmonics due to discontinuities. The value of r should be odd. This will ensure that the number of pulses for the positive half and the negative half are same and symmetric thereby avoiding even harmonics.

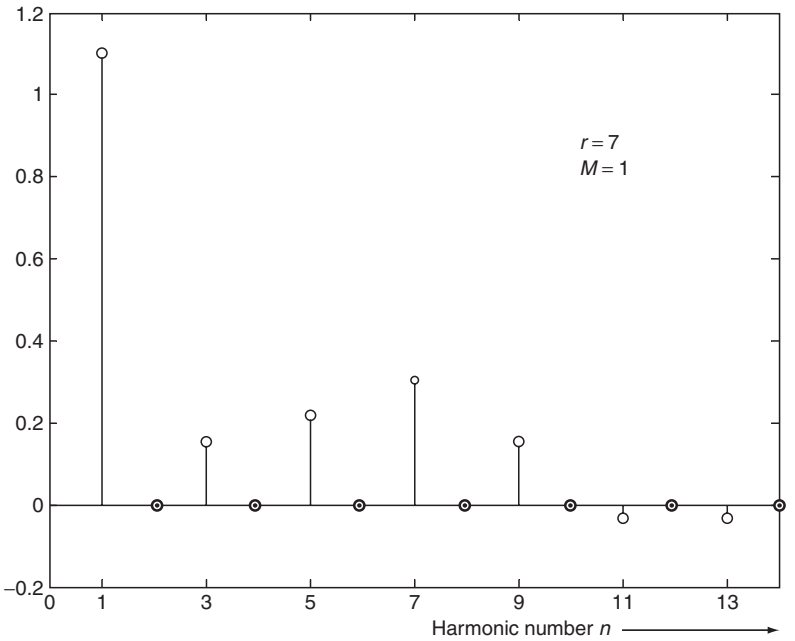


Figure 6.45 | Harmonic spectrum for unipolar PWM.

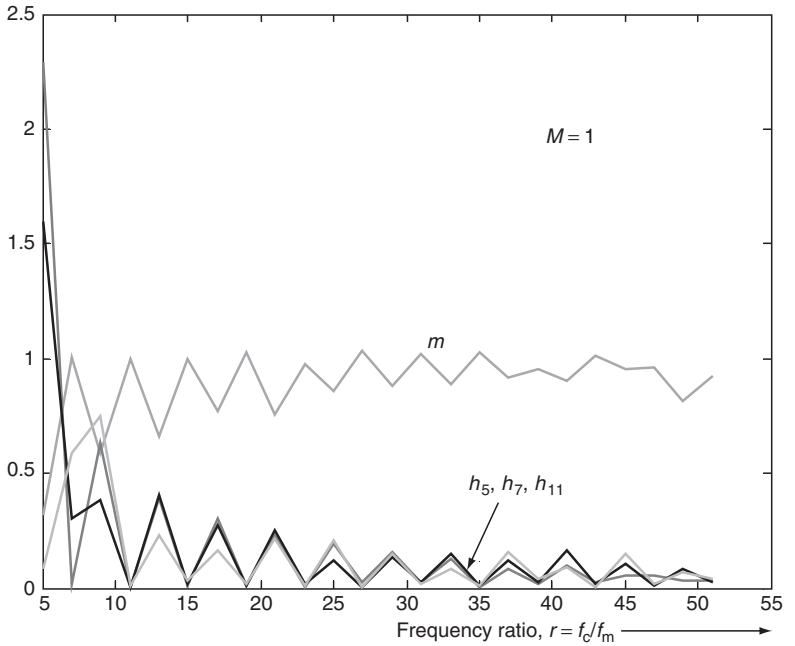


Figure 6.46 | Modulation control index and harmonic indices for bipolar PWM.

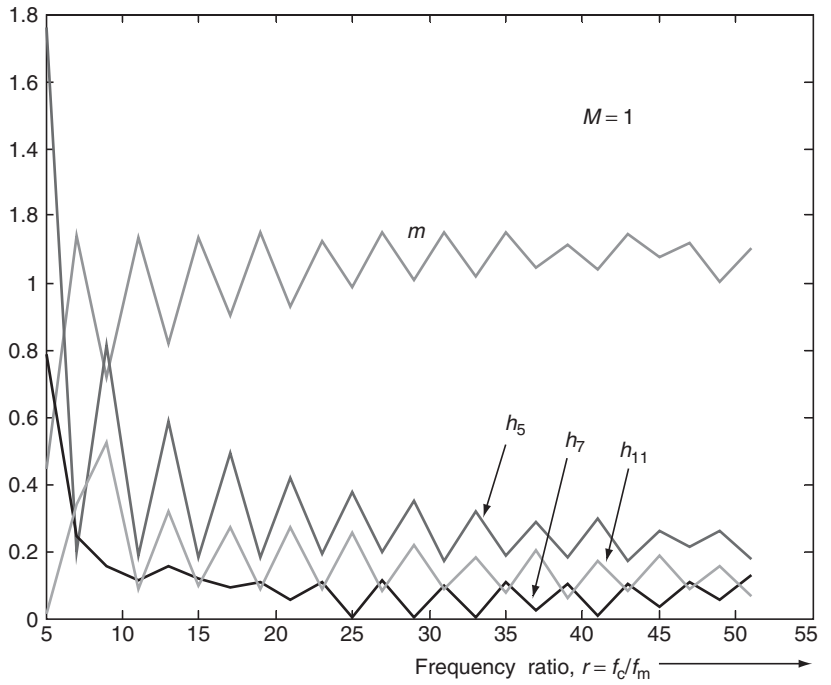


Figure 6.47 | Modulation control index and harmonic indices for unipolar PWM.

2. To ensure that the generated PWM is symmetric and does not contain even harmonics in single-phase inverters, two 180° out-of-phase sine waves are used to modulate the same triangular carrier. The pulse sequence generated by one sine wave is used to switch one arm of the full-bridge inverter and the pulse sequence generated by the out-of-phase sine wave is used to switch the other arm of the full-bridge. This will completely cancel out the even harmonics even if the frequency ratio is not odd.
3. The carrier can also be a sawtooth waveform. Either a positive slope or a negative slope sawtooth waveform can be used for modulating the sine wave. However, the notch angles must be evaluated accordingly for the sawtooth waveform to perform analysis.
4. For frequency ratios greater than $r = 20$, the relative harmonic content of the bipolar PWM reduces significantly and the integer constraint on r can be relaxed.
5. Harmonic content of the harmonic number $n = r$ is quite significant. This can be filtered out by a low-pass filter as indicated in Figure 6.48. This harmonic can be significantly reduced by using unipolar PWM.
6. In the case of three-phase systems, the PWM pattern for each arm of the inverter is displaced by 120° with respect to the other two arms. In the line-to-line voltage, the third harmonic and the harmonics that are multiples of three will be eliminated. Thus if the frequency ratio is chosen as a multiple of three, then $n = r$ will be automatically eliminated in the line voltage waveforms for three-phase loads.
7. For modulation index, $M > 1$, the notches will gradually disappear and the fundamental voltage V_1 will no longer be proportional to M. The operation when $M > 1$ is called *over-modulation*.

There are many variations from the implementation point of view. For the bipolar PWM, a triangular carrier that swings from –1 to +1 is used. However for implementing the unipolar PWM, steering and sequencing of the

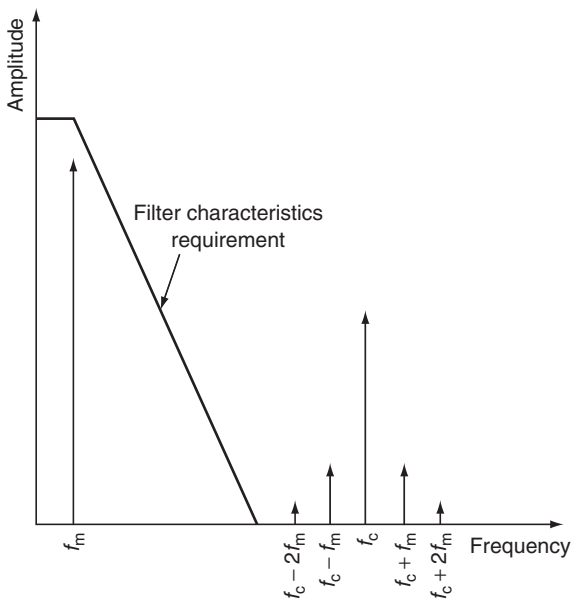


Figure 6.48 | Frequency spectrum of pulse-width modulated waveform.

inverter switching is needed. For a full-bridge inverter as depicted in Figure 6.49, the gate drive signal strategy for generating a sinusoidal PWM waveform is shown in Figure 6.50. The high-frequency PWM for every alternate half-cycle of the fundamental is given to Q_1 and Q_3 of the full-bridge inverter. Alternate half-cycle pulses at fundamental frequency (10 ms pulse for a 50 Hz sine wave) are applied to Q_2 and Q_4. The voltage across AB is a PWM voltage as shown in the waveforms of Figure 6.50. Here one should note that the flux ϕ in the transformer is given using

$$V_{AB} = N\frac{d\phi}{dt} \tag{6.50}$$

and therefore

$$\phi = \frac{1}{N}\int V_{AB}dt \tag{6.51}$$

As V_{AB} is a PWM switching waveform, it is evident from Eq. (6.51), that the flux will be the integral of the PWM waveform. The integral of the PWM waveform will have a frequency of the modulating signal and hence the flux in the transformer core will also have the modulating signal frequency, that is, 50 Hz for a 50 Hz sine wave. Therefore, the transformer has to be designed for the modulating frequency. As a consequence, the size of the transformer will become large.

In order to reduce the size of the transformer, configurations will have to be used such that the flux in the transformer core will have the frequency of the carrier and not that of the modulating signal. One of the ways to achieve high-frequency isolation is to shift the control of the output fundamental amplitude to the control of the inverter DC-link voltage. The inverter is operated at a fixed modulation index to convert the DC-link voltage to sinusoidal PWM output. The amplitude of the output is controlled by varying

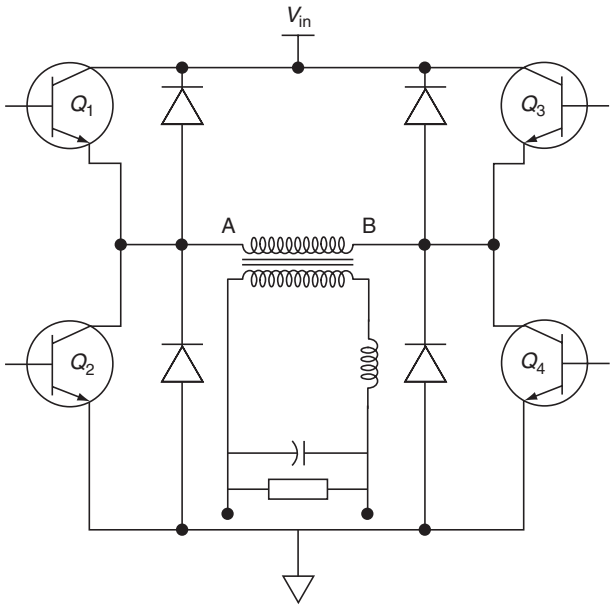


Figure 6.49 Full-bridge inverter with output filter.

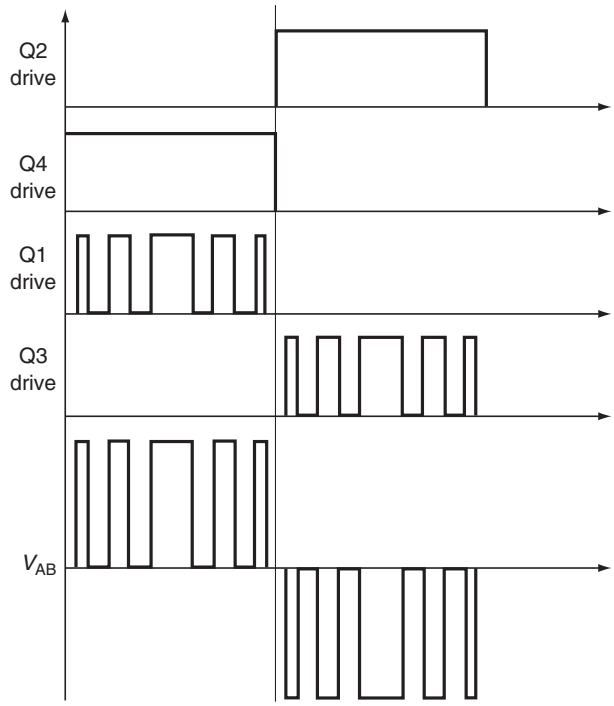


Figure 6.50 Unipolar sinusoidal pulse-width modulated waveform generation in a full-bridge inverter.

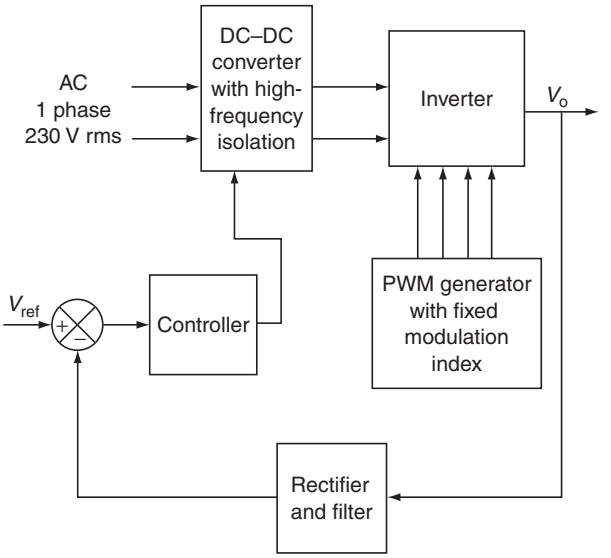


Figure 6.51 | PWM inverter schematic with high-frequency transformer for isolation.

the DC-link voltage value. The isolation transformer is within the DC–DC converter wherein the transformer can be designed for high carrier frequencies thereby reducing the size of the core. This scheme is shown in the block diagram of Figure 6.51. But one should note that here two power stages are involved (DC–DC converter and inverter) and therefore the efficiency is lower compared to the previous scheme wherein only one power stage, that is, the inverter only is used.

PWM through Phase Modulation

This is another scheme wherein the isolation transformer is a high-frequency transformer. This scheme (shown in Figure 6.52) has the following important features:

1. The transformer always sees high-frequency square wave of 50% duty cycle at the carrier frequency across its primary and hence the transformer is compact in size.
2. The filter circuit sees a PWM waveform with a carrier frequency that is twice the switching frequency. Hence the filter requirements become less stringent.

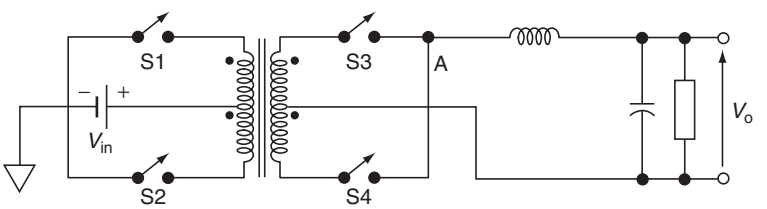


Figure 6.52 | Phase modulation inverter topology.

Principle of Operation

Let S1, S2, S3 and S4 be semiconductor switches. Switches S3 and S4 are, in particular, bi-directional switches. Switches S1 and S2 are switched complimentary to each other. Switches S3 and S4 are switched complimentary to each other. However, there exists a phase shift θ between the primary-side switching and the secondary-side switching of the transformer as shown in Figure 6.53.
From the waveforms shown in Figure 6.53, one can observe the following:

1. The voltage at A (V_A) has a switching waveshape that is twice the switching frequency, hence the output filter (L–C) design is at twice the switching frequency.
2. If the phase shift between the primary and the secondary patterns is modulated in a sinusoidal manner, then V_A is a sinusoidal PWM waveform.
3. The transformer always sees 50% duty cycle at switching frequency. Therefore, the transformer can be designed for the switching frequency, which will result in a smaller sized transformer.
4. One should note that S3 and S4 should be bi-directional switches, that is, they should allow current to flow in either direction. A bi-directional switch can be built as shown in Figure 6.54.

A bi-directional switch can be constructed using a BJT or a MOSFET or an IGBT as depicted in Figure 6.54. A current flowing from A to B follows the path A–D_1–Q–D_4–B and a current flowing from B to A follows the path B–D_3–Q–D_2–A. Note that in both the current directions, the current through the transistor switch Q is always in the same direction. Though the push–pull configuration is useful to demonstrate the principle of phase modulation, half-bridge or full-bridge configurations are usually preferred on

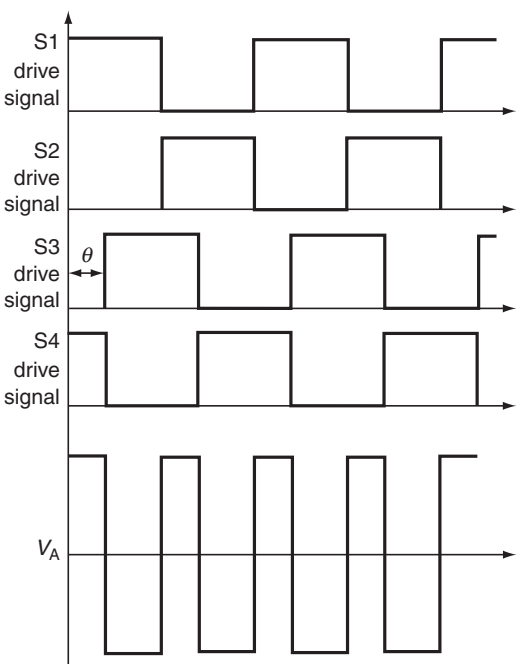


Figure 6.53 | Waveforms for the phase modulation inverter topology.

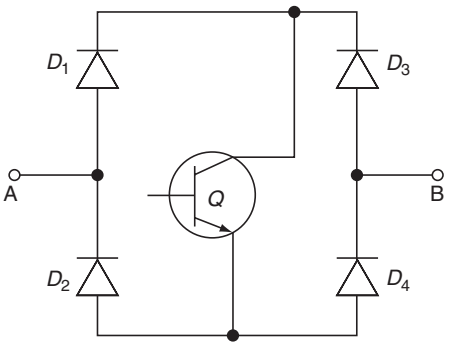


Figure 6.54 | Bi-direction switch using a BJT.

the primary side. On the secondary side, instead of a center-tapped winding, a full-bridge configuration can be used because the voltage rating for the secondary switches in the full-bridge configuration will be half that for the center-tapped winding configuration, given the same output voltage requirements.

Space-Vector Modulation

Space-vector modulation is another PWM generation principle. This technique is employed primarily for three-phase inverter systems. However this technique can be adopted for two-phase system also and extended to the poly-phase systems. Some of the features of this method that have made it a very popular three-phase PWM generation method are:

1. It is inherently suitable for digital implementation.
2. It provides a higher DC-link voltage utilization compared to the sinusoidal PWM.
3. It provides lower relative harmonic content compared to the sinusoidal PWM.
4. The switching frequency of the inverter switches is half the carrier frequency. Therefore, the switching losses are lesser as compared to a sinusoidal PWM with the same carrier frequency.

Primary Vectors

The three-phase system is described by a co-planar vector space with three axes. The three-phase inverters have three bridge arms. Each arm represents an axis in the co-planar vector space. Therefore, there are three axes, one for each arm of the bridge inverter that is uniformly distributed in the co-planar vector space at an angular distance of 120° with respect to each other. Any vector in the co-planar vector space can be represented by a triple (a_1, b_1, c_1). The co-planar vector space can be described by three primary vectors (0, 0, 1), (0, 1, 0) and (1, 0, 0). The primary vectors of the co-planar vector space of the three-phase system are depicted in Figure 6.55. Using the *a*-axis as the reference axis, the primary vectors in the polar form are represented as

$$(0,0,1) \Rightarrow 1 \cdot e^{j0} = 1$$

$$(0,1,0) \Rightarrow 1 \cdot e^{j(2\pi/3)} = e^{j(2\pi/3)}$$

$$(1,0,0) \Rightarrow 1 \cdot e^{j(4\pi/3)} = e^{j(4\pi/3)}$$

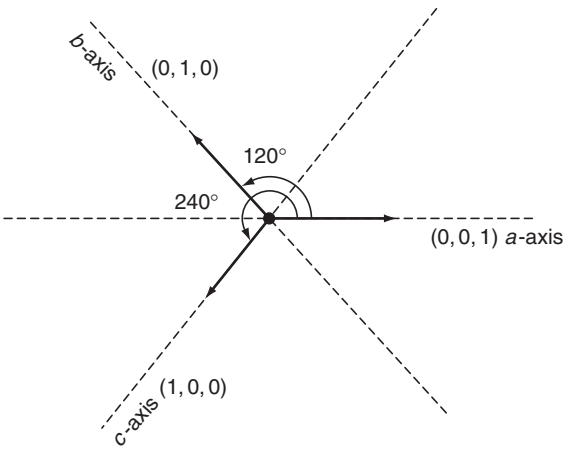


Figure 6.55 | Primary vectors defining the co-planar vector space.

Any vector in the co-planar vector space is called a *space vector* and is represented in terms of the primary vectors as

$$\vec{V} = k_1 \cdot 1 + k_2 \cdot e^{j(2\pi/3)} + k_3 \cdot e^{j(4\pi/3)} \tag{6.52}$$

where k_1, k_2 and k_3 are scalar quantities along *a*, *b* and *c* axes, respectively.

Space Vectors

Figure 6.56 shows the three primary states of the inverter bridge arms that are connected to a three-phase star load. Each arm is a two-level switch arm. When only one of the bridge arms is connected to the positive rail, then such a state is a primary state. The state (0, 0, 1) occurs when the A phase switch is connected to V_{dc} and the B and C phase switches are connected to the ground. Likewise, the state (0, 1, 0) occurs when only the B phase switch is connected to the positive rail and the state (1, 0, 0) occurs when only the C phase switch is connected to V_{dc}. The three primary states of the inverter are represented as vectors in the co-planar vector space as

$$\vec{V}_a = V_{dc} \cdot 1 + 0 \cdot e^{j(2\pi/3)} + 0 \cdot e^{j(4\pi/3)} = V_{dc} \tag{6.53}$$

Similarly

$$\vec{V}_b = V_{dc} e^{j(2\pi/3)} \tag{6.54}$$

$$\vec{V}_c = V_{dc} e^{j(4\pi/3)} \tag{6.55}$$

Each of the inverter bridge arms can assume two levels, either V_{dc} or 0. This implies that there are 2^3 states (i.e., 8 states) in all. The eight different states of the inverter including the primary states are depicted in Figure 6.57. The last two states, namely, state (0, 0, 0) and state (1, 1, 1) result in zero vectors in the co-planar

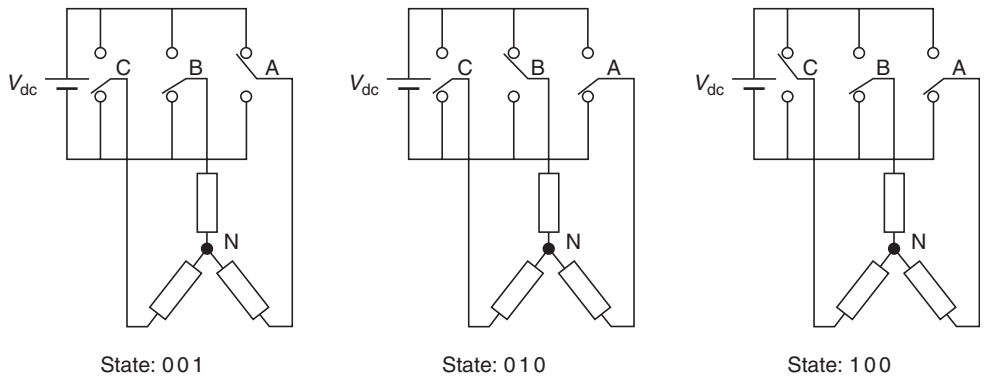


Figure 6.56 | Primary states of a two-level three-arm bridge inverter.

vector space. This means that when all the switches are connected to V_{dc} or when all the switches are connected the 0, the resulting space vector has amplitude of zero.

Figure 6.58 shows the space vectors that are obtained for the eight different states of the inverter bridge. There are two zero space vectors and they are represented as

$$V_{00} = 0 \cdot 1 + 0 \cdot e^{j(2\pi/3)} + 0 \cdot e^{j(4\pi/3)} = \mathrm{O} \Rightarrow (0,0,0)$$

$$V_{01} = 1 \cdot 1 + 1 \cdot e^{j(2\pi/3)} + 1 \cdot e^{j(4\pi/3)} = \mathrm{O} \Rightarrow (1,1,1)$$

$$V_1 = V_{dc} \Rightarrow (0,0,1) \text{ (the primary vector along } a\text{-axis)}$$

$$V_3 = V_{dc} \cdot e^{j(2\pi/3)} \Rightarrow (0,1,0) \text{ (the primary vector along } b\text{-axis)}$$

$$V_5 = V_{dc} \cdot e^{j(4\pi/3)} \Rightarrow (1,0,0) \text{ (the primary vector along } c\text{-axis)}$$

The other three vectors can be constructed from the three primary vectors. They are

$$V_2 = V_{dc} \cdot 1 + V_{dc} \cdot V_{dc} \cdot e^{j(2\pi/3)} \Rightarrow (0,1,1)$$

$$V_4 = V_{dc} \cdot e^{j(2\pi/3)} + V_{dc} \cdot V_{dc} \cdot e^{j(4\pi/3)} \Rightarrow (1,1,0)$$

$$V_6 = V_{dc} \cdot 1 + V_{dc} \cdot e^{j(4\pi/3)} \Rightarrow (1,0,1)$$

The six space vectors and the two zero space vectors form a hexagonal pattern as shown in Figure 6.58. The hexagon is divided into six sectors of 60° each. This division is made to simplify the analysis. Analysis made for any one sector can be applied to the other sectors with appropriate angular bias.

Space-Vector Modulation Principle

Consider a space vector of amplitude kV_{dc} that is aligned along the *a*-axis as shown in Figure 6.59. Let this constant amplitude space vector rotate with an angular speed of ω starting from the *a*-axis. If an observer is stationed on the *a*-axis and observes the projection of the vector tip onto an axis orthogonal to the *a*-axis and

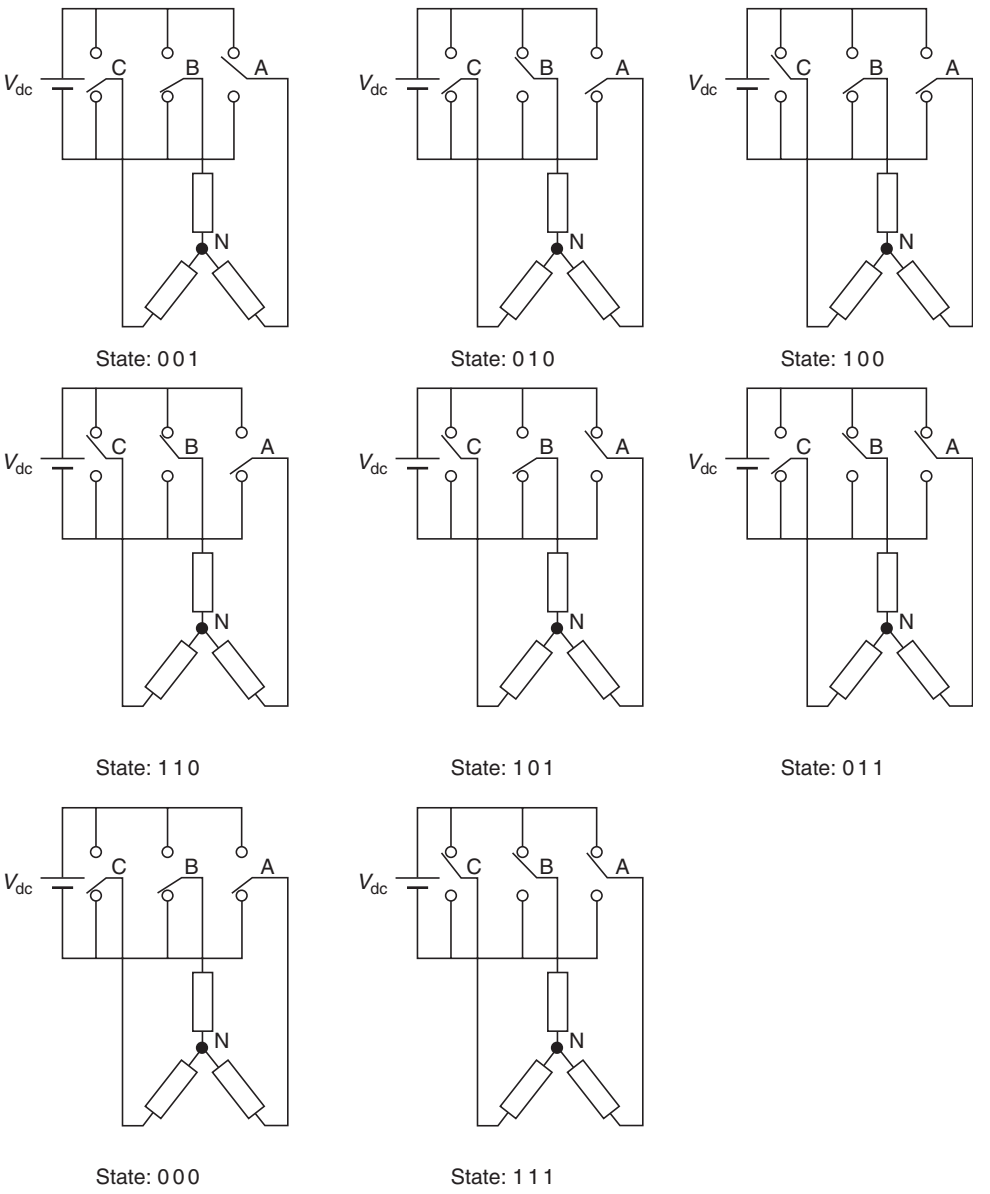


Figure 6.57 States of a two-level three-arm bridge inverter.

plots these projection as time evolves, a sinusoidal waveform with respect to time (V_{AN}) is obtained as shown and is given as

$$V_{AN} = kV_{dc} \sin \omega t \tag{6.56}$$

Similarly, if an observer is stationed on the *b*-axis and observes the projection of the vector tip onto an axis orthogonal to the *b*-axis and plots these projects as time evolves, a sinusoidal waveform with respect to time

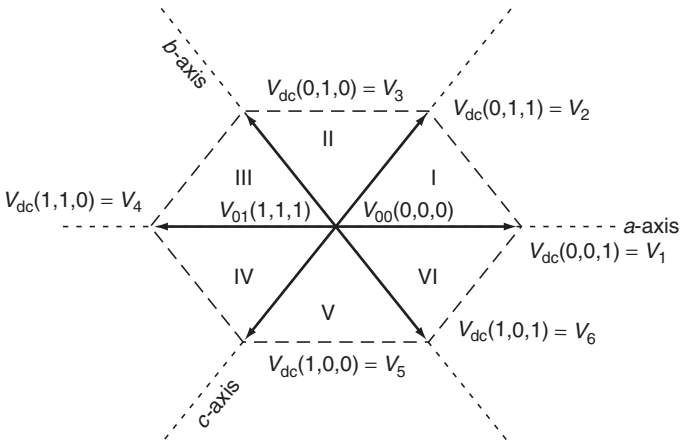


Figure 6.58 | Space vectors in the co-planar vector space.

(V_{BN}) is obtained. However, in this case the *b*-axis is positioned at an angular distance of 120° from the *a*-axis in the co-planar vector space. If the vector completes a circle in a period T, then it would take a time of $T/3$ to reach the *b*-axis. Thus the V_{BN} waveform will lag the waveform V_{AN} with respect to time or equivalently the delay in angle is

$$\omega\frac{T}{3} = \frac{2\pi}{T}\cdot\frac{T}{3} = \frac{2\pi}{3}\text{ radian}$$

Thus, V_{BN} is given as

$$V_{BN} = kV_{dc}\sin\left(\omega t - \frac{2\pi}{3}\right) \tag{6.57}$$

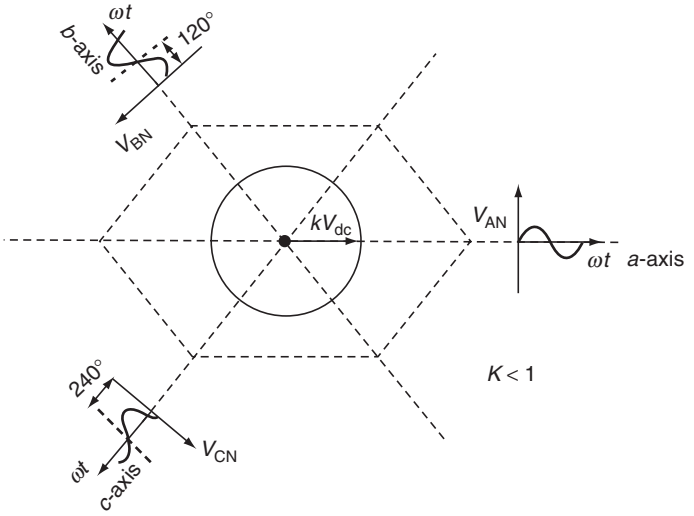


Figure 6.59 | Amplitude–time projections on the three axes for a constant amplitude rotating vector in the vector space.

Similarly, the time evolution of the projection of the space-vector tip onto an axis orthogonal to the *c*-axis is given as

$$V_{CN} = kV_{dc} \sin\left(\omega t - \frac{4\pi}{3}\right) \tag{6.58}$$

Thus one can conclude that *if a constant amplitude space vector rotates at a constant speed, then the resulting waveform with respect to time on the three axes are sinusoidal and displaced in time with respect to each other by 120°*. Conversely, one can also conclude that *if three-phase sinusoidal waveforms of frequency* ω *and displaced with respect to each other by 120° in time are applied at the three axes, then the resulting space vector is of constant amplitude and rotates at a constant speed of* ω. This is the basic principle that is used in space-vector modulation to obtain sinusoidal PWM.

The two-level three-phase inverter can produce only eight space vectors including the two zero vectors as depicted in Figure 6.58. Using combinations of these eight space vectors, any given vector within the hexagon can be obtained by time averaging in a carrier period. This means that in a carrier period T_c, the inverter spends T_1 duration of time in state (0, 0, 1) and T_2 duration of time in state (0, 1, 1) and T_0 duration of time in the zero vector state (0, 0, 0) or (1, 1, 1). Time averaging the space vectors for a carrier period of T_c gives an equivalent averaged vector kV_{dc} as shown in Figure 6.60. This equivalent averaged vector is made to rotate at a speed ω_m. Then the time evolution of the waveforms on the three axes is sinusoidal as described by Eqs. (6.56)–(6.58).

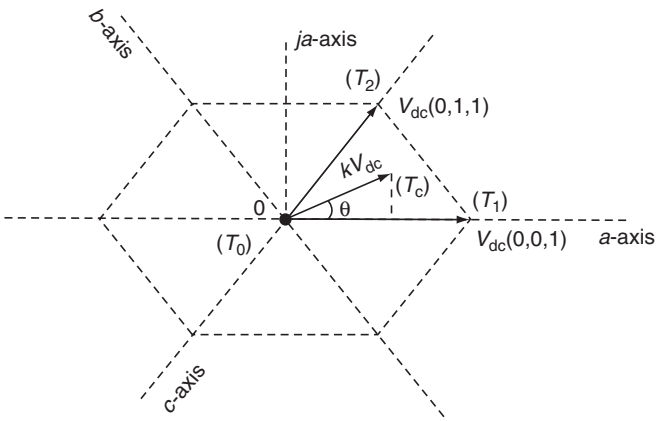


Figure 6.60 | Arbitrary space-vector realization based on time-averaged state vectors.

Consider the sector I of the hexagon as shown in Figure 6.60. If one has to realize a time-averaged space vector of amplitude kV_{dc} at an angle θ from the *a*-axis, then the inverter must spend time in the bounding states of the sector and the zero space-vector state. Let the inverter spend T_1 duration of time in state (0, 0, 1), T_2 duration of time in state (0, 1, 1) and T_0 duration of time in the zero space-vector state. Then equivalently the inverter is said to output a time-averaged space vector of amplitude kV_{dc} at an angle θ from the *a*-axis for the duration T_c, the carrier period. To derive the time durations, the time-averaged space vector can be projected onto the *a*-axis and an axis orthogonal to the *a*-axis called the *ja*-axis as shown in Figure 6.60. Referring to Figure 6.60, projecting onto the *a*-axis, one obtains

$$kV_{dc}(\cos\theta)T_c = V_{dc}T_1 + V_{dc}\cos\left(\frac{\pi}{3}\right)T_2 \tag{6.59}$$

Projecting onto the *ja*-axis, one obtains

$$kV_{dc}(\sin\theta)T_c = V_{dc}\sin\left(\frac{\pi}{3}\right)T_2 \tag{6.60}$$

From Eqs. (6.59) and (6.60), the two time durations T_1 and T_2 can be estimated. Thus, from Eq. (6.60),

$$T_2 = \frac{2}{\sqrt{3}}kT_c(\sin\theta) \tag{6.61}$$

where T_c is the carrier period and $0 \le k \le 1$. From Eqs. (6.58) and (6.60), one obtains

$$T_1 = kT_c\left[\cos\theta - \frac{1}{\sqrt{3}}\sin\theta\right]$$

$$T_1 = \frac{2}{\sqrt{3}}kT_c\left[\frac{\sqrt{3}}{2}\cos\theta - \frac{1}{2}\sin\theta\right]$$

$$T_1 = \frac{2}{\sqrt{3}}kT_c\sin\left(\frac{\pi}{3} - \theta\right) \tag{6.62}$$

The carrier period T_c is given as

$$T_c = T_1 + T_2 + T_0$$

Therefore, the time spent by the inverter in the zero vector state is

$$T_0 = T_c - T_1 - T_2 \tag{6.63}$$

For sector I, the reference state is (0, 0, 1). For sector-II, state (0, 1, 1) becomes the reference state and the Eqs. (6.61)–(6.63) are used to estimate the T_1, T_2 and T_0 times. Likewise for every sector, Eqs. (6.61)–(6.63) are used to estimate the T_1, T_2 and T_0 times. However, the bounding states are different and therefore, the system state will accordingly change depending on the sector of interest. If a space vector has an angle *a* measured from the *a*-axis, then angle *q* for the calculation of the state times depends on the sector number and is given as

$$\theta = \alpha - (s-1)\cdot\frac{\pi}{3}$$

where

$$s = \text{truncate}\left(\frac{\alpha}{\pi/3}\right) + 1$$

is the current sector within which the space vector exists. The angle θ will change at the rate of ω_m, the modulating sinusoid frequency. Equations (6.61)–(6.63) are used to estimate the times T_1, T_2 and T_0 for every carrier period interval, which in turn will determine the inverter switch states during the carrier period.

The value of *k* which scales the primary space-vector amplitude can vary from 0 to 1. However, if the space vector has to maintain a circular trajectory, then the amplitude of the space vector is limited

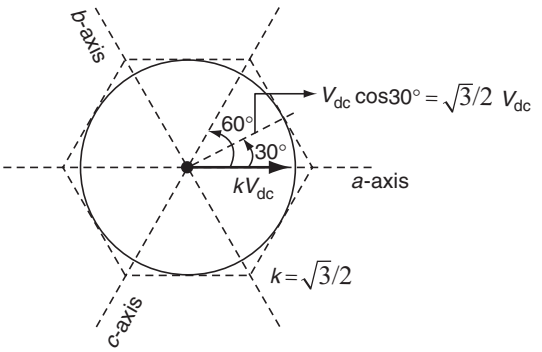


Figure 6.61 | Maximum circular space-vector trajectory.

to $(\sqrt{3}/2)V_{dc}$ such that the hexagon sides are tangents to the circular trajectory as shown in Figure 6.61. If the value of k is increased beyond $\sqrt{3}/2$ then the trajectory inscribed by the space vector will not be circular but clamped near the centers of the hexagon sides. Therefore, for $k > \sqrt{3}/2$, additional harmonics will be introduced and the inverter is said to be in over-modulation. As the value of k is increased to 1, the trajectory gradually changes over to the hexagonal trajectory as shown in Figure 6.62. In the limit, the inverter can be in each of the six states for an angular period of 60°. This will result in the classical six-step waveform as discussed in the previous section.

Sequencing

Figure 6.63 shows the timing diagram wherein during the T_1 period the PWM signal given to the three phases is such that the inverter is in state (0, 0, 1). During T_2 the inverter signals are (0, 1, 1) as indicated in the figure. During the zero vector period, either (0, 0, 0) or (1, 1, 1) can be used. However, the best harmonic and switching loss performance is obtained if T_0 period is split into two and both zero vectors are

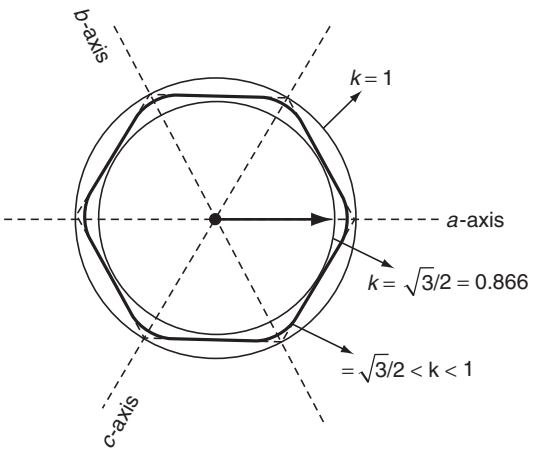


Figure 6.62 | Maximum circular space-vector trajectory for over-modulation.

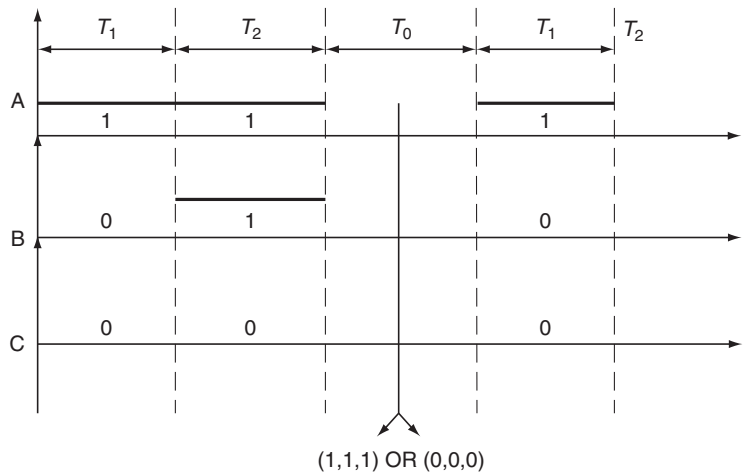


Figure 6.63 | Choice of zero vectors.

used in a carrier period. Figure 6.64 shows two carrier period cycles wherein the T_0 period is split into two $T_0/2$ periods that are appended to either side of T_1 and T_2 as shown. *The zero vector periods are chosen such that during any period only one inverter arm will switch*. This leads to the A, B and C arm switching patterns as indicated in Figure 6.64. Observe that due to the choice of the zero vectors, as indicated the switching period of the inverter, T_s is twice that of the carrier period. Therefore, *the inverter arms switch at half the carrier period, thereby reducing the switching losses* for a given carrier.

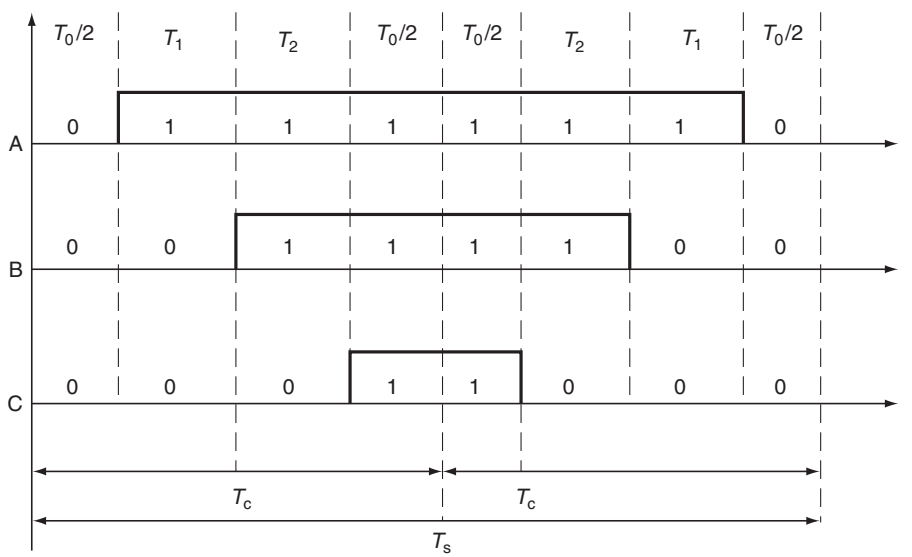


Figure 6.64 | Timing sequence.

The zero vectors are appropriately chosen for every sector of the hexagon such that the switching frequency is half of the carrier frequency. The switching sequence for two carrier periods when the time-averaged space vector is in various sectors are

Sector-I

(0,0,0)	(0,0,1)	(0,1,1)	(1,1,1)	(1,1,1)	(0,1,1)	(0,0,1)	(0,0,0)
$\frac{T_0}{2}$	T_1	T_2	$\frac{T_0}{2}$	$\frac{T_0}{2}$	T_2	T_1	$\frac{T_0}{2}$

Sector-II

(1,1,1)	(0,1,1)	(0,1,0)	(0,0,0)	(0,0,0)	(0,1,0)	(0,1,1)	(1,1,1)
$\frac{T_0}{2}$	T_1	T_2	$\frac{T_0}{2}$	$\frac{T_0}{2}$	T_2	T_1	$\frac{T_0}{2}$

Sector-III

(0,0,0)	(0,1,0)	(1,1,0)	(1,1,1)	(1,1,1)	(1,1,0)	(0,1,0)	(0,0,0)
$\frac{T_0}{2}$	T_1	T_2	$\frac{T_0}{2}$	$\frac{T_0}{2}$	T_2	T_1	$\frac{T_0}{2}$

Sector-IV

(1,1,1)	(1,1,0)	(1,0,0)	(0,0,0)	(0,0,0)	(1,0,0)	(1,1,0)	(1,1,1)
$\frac{T_0}{2}$	T_1	T_2	$\frac{T_0}{2}$	$\frac{T_0}{2}$	T_2	T_1	$\frac{T_0}{2}$

Sector-V

(0,0,0)	(1,0,0)	(1,0,1)	(1,1,1)	(1,1,1)	(1,0,1)	(1,0,0)	(0,0,0)
$\frac{T_0}{2}$	T_1	T_2	$\frac{T_0}{2}$	$\frac{T_0}{2}$	T_2	T_1	$\frac{T_0}{2}$

Sector-VI

(1,1,1)	(1,0,1)	(0,0,1)	(0,0,0)	(0,0,0)	(0,0,1)	(1,0,1)	(1,1,1)
$\frac{T_0}{2}$	T_1	T_2	$\frac{T_0}{2}$	$\frac{T_0}{2}$	T_2	T_1	$\frac{T_0}{2}$

It should be noted that the modulating signal for the space-vector modulation is a composite waveform consisting of the sinusoidal fundamental and a third harmonic component that is introduced due to splitting T_0 into two $T_0/2$ durations placed at either ends of T_1 and T_2 durations. The third harmonic component tends to flatten the top of the modulating waveform, thereby increasing the modulation *control index*. The equivalent modulating signal, compared with a triangular carrier and results in the space vector modulated output, is shown in Figure 6.65. The equivalent modulating signal consists of the dominant fundamental component f_b and a third harmonic component f_a. The equivalent modulating signal is the algebraic sum of the functions f_a and f_b.

As can be observed from Figure 6.65, the modulating signal is the summation of the sinusoidal fundamental and a third harmonic triangular component. The third harmonic triangular component flattens the top of the sinusoidal component. At a modulation index $M = 1$, the fundamental amplitude is 15% more as

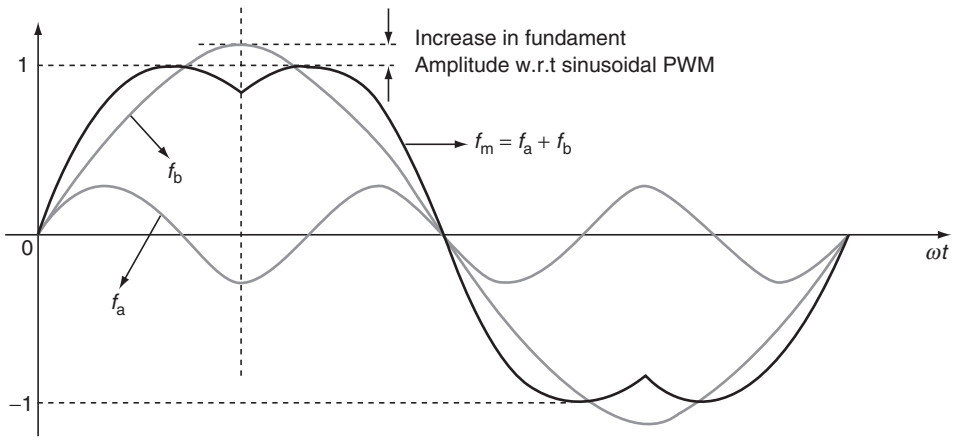


Figure 6.65 | Equivalent modulating signal of space-vector PWM.

compared to that of the sinusoidal PWM. In sinusoidal PWM, the control index $m = 1$ compared to $4/\pi$ for the unmodulated waveform. This is about 78.5% of the unmodulated waveform. In the case of the space-vector PWM, the fundamental amplitude is 93% of the unmodulated waveform.

Multi-Level Inverters

The space-vector modulation concept can be extended to multi-level inverters too. Figure 6.66 shows the space vectors for the four-level three-arm bridge converter.

As discussed for the two-level inverter, the co-planar vector space can be described by three primary vectors (0, 0, 1), (0, 1, 0) and (1, 0, 0). Any vector in the co-planar vector space (i.e., space vector) is represented in terms of the primary vectors. Each of the axes consists of four levels: 0, $k_1 V_{dc}$, $k_2 V_{dc}$ and V_{dc}. Thus there are four primary space vectors for each of the three axes. These primary vectors can be vectorially added to obtain the other states of the multi-level inverter. For sector-I, the state vectors along the *a*-axis and the negative *c*-axis can be combined to form three more space-vector states as shown. Thus, each sector gets divided into nine smaller sub-sectors.

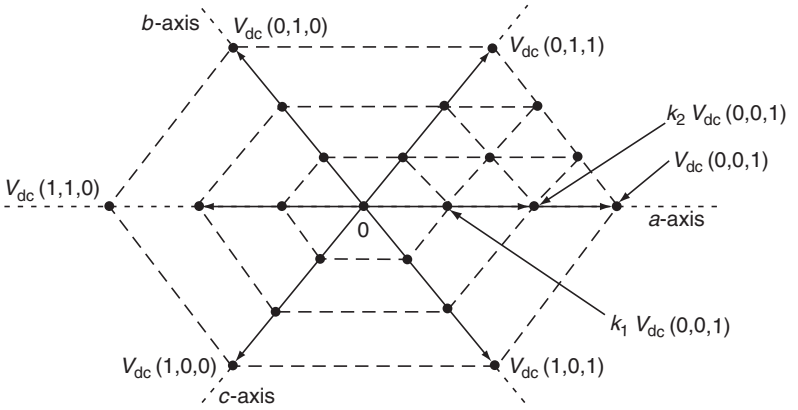


Figure 6.66 | State vectors in the co-planar vector space for four-level three-arm bridge.

As there are more states available, there is more flexibility in obtaining the equivalent time-averaged rotating space vector. Different choice of states and sequencing can lead to different optimizations. As there are more states available for the inverter, the carrier frequency to the fundamental modulating frequency ratio (r) can be significantly reduced.

6.7 Pulse–Width Modulation with Harmonic Elimination

There are many applications wherein harmonics are detrimental to the functioning of the system. In applications involving transformers and AC motors, the harmonics will lead to excess core losses and consequently heat up the core material which leads to deterioration of the magnetic core. Further, in some applications some harmonics lead to undesirable effects. In the case of single-phase AC motors like fans, the harmonics can lead to torque pulsations and audible noise generation. Therefore to reduce the adverse effects sometimes few specific undesirable harmonics like third, fifth, seventh, etc. need to be eliminated.

Consider the quasi-square waveform as shown in Figure 6.67. The fundamental component is given by setting $n = 1$. This is given as

$$f_{\text{fundamental}} = \frac{4V}{\pi}\cos\theta \sin \omega t \tag{6.64}$$

The control is performed by duty cycle control as this approach gives a greater control over the harmonic elimination rather than the DC-link control. Using this approach, from Eq. (6.64), it can be observed that there is only one variable that can be controlled and that is θ. With this, one can either

control the amplitude of the fundamental only

OR

control the amplitude of one particular harmonic

OR

eliminate one particular harmonic

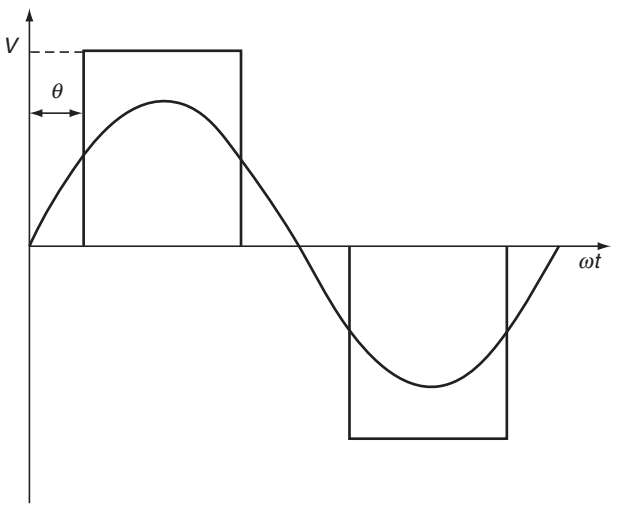


Figure 6.67 | Quasi-square waveform and its fundamental.

For example, if one desires to eliminate the third harmonic, then one has to set

$$\frac{4V}{\pi}\cdot\frac{1}{3}\cdot\cos(3\theta)\sin(3\omega t)=0$$

which implies that $3\theta = 90^{o}$ or $\theta = 30^{o}$. Therefore, if one sets $\theta = 30^{o}$ for the quasi-square waveform, then the third harmonic gets eliminated. Thus, one can generalize that by setting $\theta = 90^{o}/n$, the nth harmonic can be eliminated. If one also needs to control the amplitude of the fundamental in addition to eliminating one harmonic, then two control variables are needed, one to eliminate the harmonic and one to set the fundamental to the desired level.

Harmonic Elimination with Multiple Notches

With a quasi-square waveform, there is only one control variable and hence either the fundamental can be controlled or a harmonic can be eliminated. However, if one needs to control many parameters of the output waveform then a more complex waveform structure is needed. Consider the waveform shown in Figure 6.68 that contains two control variables, θ_1 and θ_2. Note that the waveform has quarter-wave symmetry. This implies that the Fourier series would then contain only "sine" terms. θ_1 and θ_2 are called the "NOTCH" angles. At every notch, the voltage transits from either positive to negative or negative to positive resulting in voltage reversal. If the "notches" θ_1 and θ_2 are specified, then the whole waveshape becomes fully determined.

For the waveform of Figure 6.68 with two notches θ_1 and θ_2, the average value in a period is zero and therefore

$$a_0 = 0$$

The waveform of Figure 6.68 has quarter wave symmetry and therefore,

$$a_n = 0$$

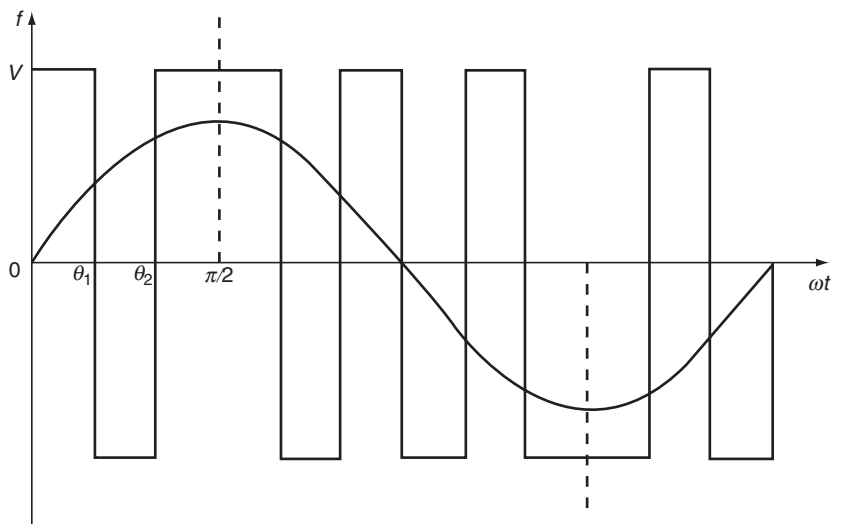


Figure 6.68 | Two-notch pulse-width modulated waveform.

Further as the waveform of Figure 6.68 is symmetrical about the *x*-axis, it contains only odd harmonics. Thus for a quarter wave symmetrical waveform containing only odd harmonics,

$$b_n = \frac{2}{\pi/2}\int_0^{\pi/2} f(\omega t)\sin(n\omega t)\mathrm{d}\omega t$$

$$b_n = \frac{4V}{n\pi}\left[1+2(-\cos n\theta_1+\cos n\theta_2)\right] \text{ for } n = 1, 3, 5, \ldots \tag{6.65}$$

For example, if one desires to set the fundamental amplitude to "x" and also to eliminate the third harmonic for the two-notch PWM waveform, then it is required that

$$b_1 = x = \frac{4V}{\pi}[1+2(-\cos\theta_1+\cos\theta_2)]$$

$$b_3 = 0 = \frac{4V}{3\pi}[1+2(-\cos 3\theta_1+\cos 3\theta_2)]$$

$$f_1 = \frac{4V}{\pi}[1+2(-\cos\theta_1+\cos\theta_2)] - x = 0 \tag{6.66}$$

$$f_2 = \frac{4V}{3\pi}[1+2(-\cos 3\theta_1+\cos 3\theta_2)] = 0 \tag{6.67}$$

Solving Eqs. (6.66) and (6.67) for the two unknowns θ_1 and θ_2, one obtains θ_1 and θ_2 such that the two-notch waveform has a fundamental amplitude of "x" and the third harmonic is absent. This principle can be extended to waveshapes having more than two notches. Let there be "p" notches per quarter cycle, then,

"p" harmonics can be eliminated

OR

"$(p-1)$" harmonics can be eliminated with a specified value for the fundamental amplitude

For "p" notches per quarter cycle with quarter wave symmetry,

$$b_n = \frac{4V}{n\pi}\left[1+2\sum_{k=1}^{p}(-1)^k\cos n\theta_k\right] \tag{6.68}$$

where θ_k is the kth notch angle. From Eq. (6.68), one obtains "p" equations which will allow θ_1, θ_2, …, θ_p to be estimated for the specified constraints on the selected harmonics.

Determination of Notch Angles

This section summarizes the harmonic elimination process by providing a step-by-step approach to determine the notch angles for specific harmonic constraints.

Step 1: Decide on the number of notches required per quarter cycle. This can be specified based on the number of harmonics one needs to eliminate and whether fundamental amplitude requires to be set to a desired value or not.

Step 2: Using Eq. (6.68), obtain the p equations in p unknown notch angles θ_1, θ_2, …, θ_p.

Step 3: The p equations obtained from Step (2) form a set of non-linear equations. The solution to the set of p equations is based on the extension of the Newton–Cotes algorithm to non-linear equations. The solution is an iterative algorithm that is given as

$$\begin{bmatrix} \theta_1 \\ . \\ . \\ \theta_p \end{bmatrix}_{k+1} = \begin{bmatrix} \theta_1 \\ . \\ . \\ \theta_p \end{bmatrix}_k - [J]^{-1} \begin{bmatrix} f_1(\theta_1,...,\theta_p) \\ . \\ . \\ f_p(\theta_1,...,\theta_p) \end{bmatrix}_k \tag{6.69}$$

where J is the Jacobian matrix for the solution of non-linear equations and is given by

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1}(\theta_1,...,\theta_p) & . & . & \frac{\partial f_1}{\partial \theta_p}(\theta_1,...,\theta_p) \\ . & & & . \\ . & & & . \\ \frac{\partial f_p}{\partial \theta_1}(\theta_1,...,\theta_p) & . & . & \frac{\partial f_p}{\partial \theta_p}(\theta_1,...,\theta_p) \end{bmatrix}_k \tag{6.70}$$

Step 4: Start the iterative algorithm by making an initial guess for the p unknown notch angles, that is

$$\begin{bmatrix} \theta_1 \\ . \\ . \\ \theta_p \end{bmatrix}_{k=0} = \begin{bmatrix} a_1 \\ . \\ . \\ a_p \end{bmatrix}$$

where a_1 to a_p are the initial guess values for θ_1 to θ_p, respectively.

Step 5: Iteratively evaluate

$$\begin{bmatrix} \theta_1 \\ . \\ . \\ \theta_p \end{bmatrix}_{k+1}$$

till the function matrix

$$\begin{bmatrix} f_1(\theta_1,...,\theta_p) \\ . \\ . \\ f_p(\theta_1,...,\theta_p) \end{bmatrix}_k < \begin{bmatrix} \varepsilon_1 \\ . \\ . \\ \varepsilon_p \end{bmatrix}$$

where ε_i are very small values that are used to break the iteration loop once the error in the function matrix value at instant k are less than the ε_I values.

EXAMPLE 6.1

Let it be required that the fifth and the seventh harmonics be eliminated from a PWM waveform.

Step 1: It is required that two harmonics are to be eliminated and there is no constraint on the amplitude of the fundamental. This implies that two notches per quarter cycle are sufficient. Hence, $p = 2$.

Step 2: Using Eq. (6.68), one obtains

$$f_1 = \cos(5\theta_1) - \cos(5\theta_2) - 0.5 = 0$$
$$f_2 = \cos(7\theta_1) - \cos(7\theta_2) - 0.5 = 0$$

Step 3: The Jacobian J is obtained as

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} \\ \frac{\partial f_2}{\partial \theta_1} & \frac{\partial f_2}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} -5\sin(5\theta_1) & 5\sin(5\theta_2) \\ -7\sin(7\theta_1) & 7\sin(7\theta_2) \end{bmatrix}$$

Formulate the iterative equation to solve for θ_1 and θ_2. It is given as

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{k+1} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_k - [J]^{-1} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}_k \tag{6.71}$$

Step 4: To start the iterative solution as given by Eq. (6.71), initial values of θ_1 and θ_2 are required. Therefore, make an initial guess for θ_1 and θ_2 at $k = 0$. Evaluate the Jacobian matrix J and $\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$ with values of θ_1 and θ_2 at $k = 0$. Using Eq. (6.71), evaluate

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{k=1}$$

Step 5: Iteratively evaluate

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{k+1}$$

till each element of

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}_k$$

is less than a predetermined value, say less than 0.0001. Then the corresponding values of θ_1 and θ_2 at iteration level $k + 1$ will be the required notch angles θ_1 and θ_2.

Implementation: One of the simple ways to implement this harmonic elimination technique is to store the values of $[\theta_1, \ldots, \theta_p]$ in a read only memory (ROM). The notch angles can be loaded into a compare register sequentially and cyclically to generate the "p" notches per quarter wave waveshape. Due to the computation involved in calculating the Jacobian and solving the non-linear equation, the harmonic elimination method is normally used with pre-computed values for specific harmonics and fundamental frequency and stored in

the ROM. However today with high compute power and low-cost digital signal processors available, it is possible to dynamically estimate the notch angles for various harmonics and specific fundamental frequency on-the-fly. Further, for the three-phase loads the computational overhead is much reduced as the third harmonic and the harmonics that are multiples of three are anyway eliminated. This means that the dominant harmonics that need to be eliminated are the 5th, 7th, 11th and 13th harmonics. This implies that only four notches are needed and a 4 × 4 Jacobian matrix inverse needs to be evaluated.

Staircase PWM

Another harmonic elimination method is the staircase PWM technique. This method combines the principles of the sinusoidal PWM and the harmonic elimination with notch method to give a simple PWM technique with higher modulation control index than the bipolar PWM. The carrier is a triangular waveform that transits between –1 and +1 like in the case of the sinusoidal PWM carrier. The modulation signal is instead a staircase waveform. The levels of the staircase waveform are decided based on the harmonics to be eliminated in a method similar to the notch-based harmonic elimination method. Figure 6.69 depicts the triangular carrier and a two-level staircase waveform superimposed on the carrier wherein the modulation index M is less than 1. The angles corresponding to the points of intersection are the notch angles at which the voltage transition takes place.

The carrier frequency or the frequency of the switching pulses is

$$\omega_c = 2\pi f_c; \quad T_c = \frac{1}{f_c}$$

and the modulating frequency or the frequency of the fundamental is

$$\omega_m = 2\pi f_m; \quad T_m = \frac{1}{f_m}$$

As discussed for the sinusoidal PWM, the ratio of the carrier frequency to the modulating frequency is called the frequency ratio r that is given as

$$r = \frac{\omega_c}{\omega_m} = \frac{f_c}{f_m} \tag{6.72}$$

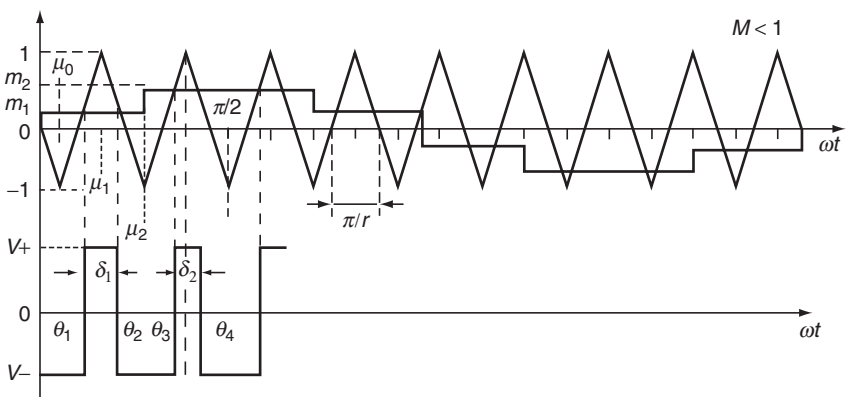


Figure 6.69 | Staircase pulse-width modulation.

Referring to Figure 6.69, the Fourier representation of the staircase PWM for "p" notches in a quarter cycle is given as

$$f_{\text{staircase}} = \sum_{n=1}^{\infty} b_n \sin(n\omega t) \tag{6.73}$$

where

$$b_n = \frac{2}{\pi/2}\int_0^{\theta_1}(-V)\sin(n\omega t)\mathrm{d}\omega t + \frac{2}{\pi/2}\int_{\theta_1}^{\theta_2} V\sin(n\omega t)\mathrm{d}\omega t + \cdots + \frac{2}{\pi/2}\int_{\theta_p}^{\pi/2}(-1)^{p+1}V\sin(n\omega t)\mathrm{d}\omega t$$

$$b_n = \frac{4V}{n\pi}\sum_{\mathrm{i}=0}^{p}\cos n\theta_{\mathrm{i}} - \cos n\theta_{\mathrm{i}+1} + \frac{4(-V)}{n\pi}\sum_{\mathrm{j}=1}^{p}\cos n\theta_{\mathrm{j}} - \cos n\theta_{\mathrm{j}+1} \tag{6.74}$$

for i = even (0, 2, 4, 6, 8, …); j = odd (1, 3, 5, 7, …); n = odd (1, 3, 5, 7, …).

$$b_n = \frac{4V}{n\pi}[\cos n0 - 2\cos n\theta_1 + 2\cos n\theta_2 - 2\cos n\theta_3 + \cdots + (-1)^p 2\cos n\theta_{\mathrm{p}}]$$

$$b_n = \frac{4V}{n\pi}[1 - 2\cos n\theta_1 + 2\cos n\theta_2 - 2\cos n\theta_3 + \cdots + (-1)^p 2\cos n\theta_{\mathrm{p}}]$$

$$b_n = \frac{-8V}{n\pi}[\cos n\theta_1 - \cos n\theta_2 + \cos n\theta_3 - \cos n\theta_4 + \cdots + (-1)^{p+1}\cos n\theta_{\mathrm{p}} - 0.5]$$

$$b_n = \frac{-8V}{n\pi}\left[2\sin\left(n\frac{\theta_1+\theta_2}{2}\right)\sin\left(n\frac{\theta_2-\theta_1}{2}\right) + 2\sin\left(n\frac{\theta_3+\theta_4}{2}\right)\sin\left(n\frac{\theta_4-\theta_3}{2}\right) + \cdots + (-)0.5\right]$$

$$b_n = \frac{-16V}{n\pi}\left[\sin\left(n\frac{\theta_1+\theta_2}{2}\right)\sin\left(n\frac{\theta_2-\theta_1}{2}\right) + \sin\left(n\frac{\theta_3+\theta_4}{2}\right)\sin\left(n\frac{\theta_4-\theta_3}{2}\right) + \cdots + (-)0.25\right] \tag{6.75}$$

For a specific level of the staircase waveform, the notch width will be the same at any comparison point with the triangular carrier. Thus for a given level, the notch width can be defined as follows:

1. **Level-1:** Notch angle width is $\delta_1 = \theta_2 - \theta_1$.
2. **Level-2:** Notch angle width is $\delta_2 = \theta_4 - \theta_3$ and so on.

The mid-points of the pulse widths are given as

$$\mu_1 = \frac{\theta_1+\theta_2}{2};\ \mu_2 = \frac{\theta_2+\theta_3}{2};\ \cdots$$

The Fourier series coefficient b_n for the staircase waveform as given in Eq. (6.75) can be re-written as

$$b_n = \frac{-16V}{n\pi}\cdot\left[\sum_{i=1}^{k_1}\sin n\mu_i \sin\left(n\frac{\delta_1}{2}\right) + \sum_{i=k_1+1}^{k_1+k_2}\sin n\mu_i \sin\left(n\frac{\delta_2}{2}\right) + \cdots + (-)0.25\right] \tag{6.76}$$

where k_1 is the number of pulses resulting from level-1 staircase intersection with the carrier; k_2 the number of pulses resulting from level-2 staircase intersection with the carrier and so on. For $k_1 = k_2 = 1$, the Eq. (6.76) reduces to

$$b_n = \frac{-16V}{n\pi}\left[\sin n\mu_1 \sin\left(n\frac{\delta_1}{2}\right) + \sin n\mu_2 \sin\left(n\frac{\delta_2}{2}\right) + \cdots + (-)0.25\right]$$

For an m-level staircase waveform, at $M = 1$, the mth level will be at par with the peak amplitude of the carrier and therefore, there will be no notches corresponding to the mth level. In a 3-level staircase waveform, there will be notches corresponding to only two levels for $M = 1$. To eliminate the fifth and the seventh harmonics, from Eq. (6.76), one has to set

$$f_1 = \left(\sum_{i=1}^{k_1} \sin 5\mu_i\right)\sin 5\frac{\delta_1}{2} + \left(\sum_{i=k_1+1}^{k_1+k_2} \sin 5\mu_i\right)\sin 5\frac{\delta_2}{2} - 0.25 = 0 \tag{6.77}$$

$$f_2 = \left(\sum_{i=1}^{k_1} \sin 7\mu_i\right)\sin 7\frac{\delta_1}{2} + \left(\sum_{i=k_1+1}^{k_1+k_2} \sin 7\mu_i\right)\sin 7\frac{\delta_2}{2} - 0.25 = 0 \tag{6.78}$$

For $k_1 = k_2 = 1$, Eqs. (6.77) and (6.78) reduce to

$$f_1 = \sin(5\mu_1)\sin\left(5\frac{\delta_1}{2}\right) + \sin(5\mu_2)\sin\left(5\frac{\delta_2}{2}\right) - 0.25 = 0 \tag{6.79a}$$

$$f_2 = \sin(7\mu_1)\sin\left(7\frac{\delta_1}{2}\right) + \sin(7\mu_2)\sin\left(7\frac{\delta_2}{2}\right) - 0.25 = 0 \tag{6.79a}$$

The midpoints of the pulse widths μ_1 and μ_2 are the center points of the triangular carrier and therefore they are known quantities:

$$\mu_i = \frac{\pi/2}{r} + i\frac{\pi}{r} = (2i+1)\frac{\pi}{2r} \quad \text{for } i = 0, 1, 2, 3, \ldots \tag{6.80}$$

Here δ_1 and δ_2 are the unknown quantities that are estimated by solving Eqs. (6.77) and (6.79) using the method outlined in section on "Determination of Notch Angles" for the harmonic elimination with multiple notches.

The levels of the staircase waveforms are proportional to the pulse widths δ_i. If the level is at the zero axis, then the intersection points with the triangular carrier are on the zero axis and the pulse width is equal to π/r. If the level is at +1, then the intersection with the carrier is at a point on the peak of the triangle and the pulse width is zero. As the carrier is a linearly rising waveform, the relationship between the pulse width and staircase level for the modulation index is illustrated in Figure 6.70. The height of the ith level is

$$m_i = M\left(1 - \frac{\delta_i}{\pi/r}\right) \tag{6.81}$$

where m_i is the height of the ith level of the staircase PWM waveform.

Summary of Staircase PWM Design

1. The fundamental component of the staircase PWM waveform has a *control index m* that is higher than the sinusoidal PWM for a given *modulation index M*.

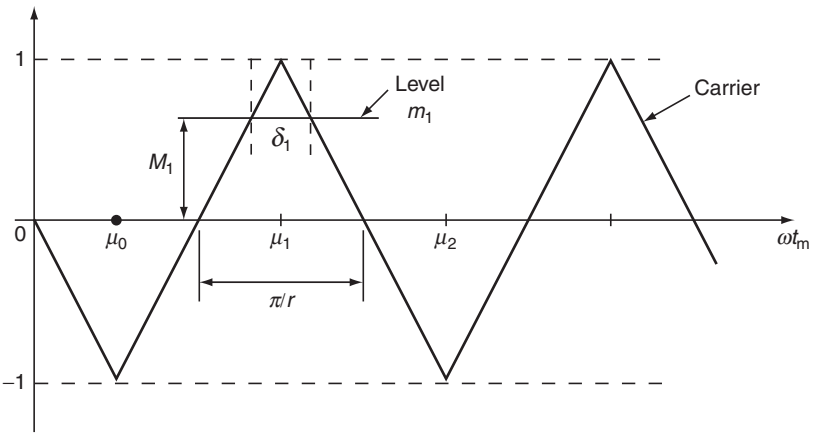


Figure 6.70 | Triangular carrier along with the pulse width and level relationship.

2. For an n-level staircase PWM, $n-1$ harmonics can be eliminated at a modulation index of $M = 1$. For modulation indices less than 1, the harmonics that are designed to be eliminated at $M = 1$ will remain eliminated or negligible.
3. To design the staircase PWM pattern for eliminating specific harmonics, first choose the number of levels for the staircase PWM.
4. Then choose the frequency ratio r.
5. Then choose the number of pulses needed for every level, that is, k_1, k_2, ….
6. Evaluate the pulse-width center points μ_1, μ_2, etc. as given by Eq. (6.80).
7. Formulate the $n-1$ equations for eliminating the $n-1$ harmonics along similar lines like Eqs. (6.78) and (6.79).
8. Solve the $n-1$ non-linear equations as discussed in the previous sub-section for harmonic elimination with multiple notches and obtain the required pulse widths for the staircase PWM.
9. From the estimated values of the pulse widths, calculate the heights of the levels of the staircase PWM.

CONCLUDING REMARKS

The vast expanse of DC–AC converters or inverters is rather difficult to discuss in a single chapter. However, the essentials have been discussed in this chapter that will give a flavor of the various issues and the insights gained will help in pursuing advanced inverter concepts. One of the most significant issues that needs to be assimilated is the concept of pulse-width modulation (PWM). It is the PWM that eventually gives one the waveform that one desires at the output. Therefore, time spent on the study of PWM is considered to be well spent.

Most of the pulse-width modulators were based on operational amplifier (op-amp) based circuits. However, today most if not all algorithms are embedded into the digital processors. This chapter does not discuss the algorithmic implementation of the PWM, but focuses more on the Fourier analysis aspects. This will provide a sound base for the design

of the PWM patterns. The algorithmic implementation aspect is dealt in Chapter 12 after a formal discussion on the discrete computation issues.

Harmonic elimination methods that are discussed in this chapter were implemented previously by lookup table methods where the patterns were pre-computed for various fundamental amplitudes and stored in ROM or EPROM. Depending on the control input, the specific pattern for the required fundamental was read out from the stored memory location. However, today, with even low-priced DSPs having good computational speeds, on-the-fly calculation of the required PWM pattern (notches or staircase levels) for harmonic elimination is a feasible possibility.

The laboratory exercises suggested in the next section are mainly for the beginners to get a good understanding of the basics of PWM and dead time delay aspects for single-phase inverters. This is in no way an exhaustive list of experiments. Based on the topics covered in the chapter, one may extend to three-phase and multi-level inverter experiments along similar lines as suggested for single-phase inverter. It should be noted that the more popular PWM methods like the space-vector PWM are not easy to implement with analog devices like the op-amp. They are best implemented as embedded algorithms on a digital processor as discussed in Chapter 12.

LABORATORY EXERCISES

1. Consider the block schematic of the pulse-width modulator as shown in Figure 6.71. It consists of a comparator wherein the modulating signal is one input and it is compared with the carrier signal connected to the other input. The output of the comparator is a PWM signal. The PWM signal is passed through a low-pass filter to obtain the filtered output.

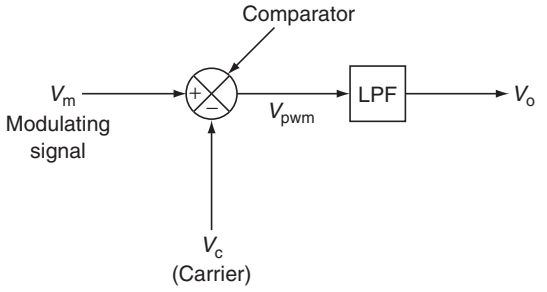


Figure 6.71 Pulse-width modulator.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Simulation in SciLAB
- **c.** Hardware bread-boarding

Tasks for study:

- **(a)** Based on the block schematic of Figure 6.71, rig up the pulse-width modulator circuit using a Schmitt comparator, one signal generator for carrier signal and another signal generator for the modulating signal. The low-pass filter can be any standard op-amp based filter like a Sallen-Key filter.
- **(b)** Set the carrier signal to be a triangular waveform swinging from –1 V to +1 V with a carrier frequency of 20 kHz. Set the low-pass filter cut-off at 2 kHz. Set the modulating signal generator to be a sinusoidal waveform of frequency 50 Hz with the amplitude at 0.5 V peak.
- **(c)** Observe the PWM output V_{pwm} and the filtered output V_o. Compare V_o and V_m and observe the differences.
- **(d)** For hardware experiment, the data may be copied to a computer in ASCII format. The data may then be imported to SciLAB or MATLAB. Check and compare the line spectrum of the harmonic content in V_m, V_{pwm} and V_o.

(e) Change the amplitude of the modulating signal and repeat Steps (3) and (4).

(f) Keeping the carrier frequency at 20 kHz and without changing the low-pass filter cut-off, increase the modulating signal frequency gradually and observe the effect on V_{pwm} and V_o. Observe the effect on the harmonic contents of V_{pwm} and V_o also.

(g) Using V_{pwm} as input, rig up a gating or steering circuit for the four switches of a full-bridge inverter.

2. Consider the block schematic of the pulse-width modulator as shown in Figure 6.72. It consists of a TL494 IC with the modulating signal given as the DTC pin of the IC. V_{pwm} and $\overline{V_{pwm}}$ are the two out-of-phase PWM signals that are required to be given to the inverter switches.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice

b. Hardware bread-boarding

Tasks for study:

(a) Rig up the pulse-width modulator circuit as shown in Figure 6.72.

(b) R_T and C_T are used to set the carrier frequency. In this case the carrier is a sawtooth waveform swinging from 0 to 3 V. Therefore, the modulating signal that is given to the DTC pin should have a bias at 1.5 V and the sinusoidal waveform with a peak value less than 1.5 V should be superimposed on the bias such that a level-shifted sine wave is obtained.

(c) Observe the PWM output, V_{pwm} and $\overline{V_{pwm}}$.

3. The top switch and the bottom switch of an inverter arm have finite rise and fall times. To ensure that the turning-OFF device has sufficient time to turn OFF, the turn-ON signal for the other switch of the arm is delayed by a few microseconds. This way a dead time is ensured at every switch transition that prevents the DC link from shorting to the ground. The dead time delay is incorporated in a digital processor by using a low bit counter. In hardware implementation, one of the ways to implement dead time delay is as shown in Figure 6.73. It consists of Schmitt AND gates connected as shown. The *RC* is connected to provide the delay for every rising edge.

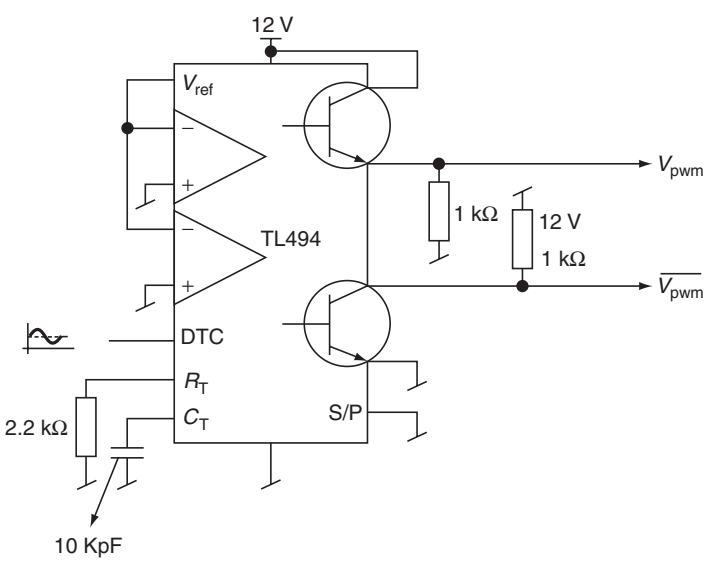


Figure 6.72 | Pulse-width modulator based on TL494 IC.

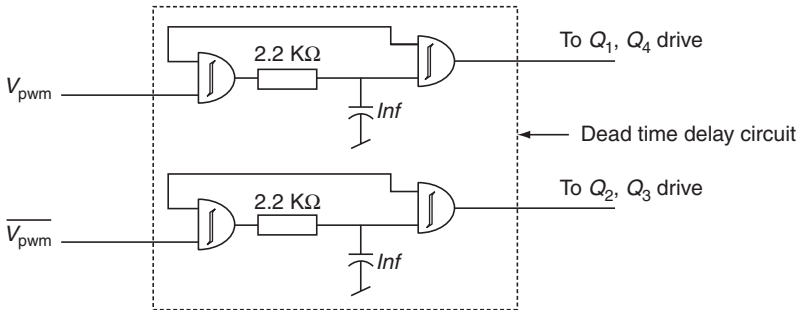


Figure 6.73 | Dead time delay circuit.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware bread-boarding

Tasks for study:

(a) Rig up the pulse-width modulator circuit followed by the dead time logic as shown in Figure 6.73.
(b) The PWM outputs, V_{pwm} and $\overline{V_{pwm}}$ are given as input to the dead time logic.
(c) Observe Q_1 and Q_2 drive signals on the oscilloscope both during the rising and falling transitions and note the dead time.
(d) Change the *RC* time constant and observe the effect on the dead time.
(e) Likewise observe the dead time between the Q_3 and Q_4 drive signals.

4. The schematic for the phase-modulated PWM is shown in Figure 6.74. The TL494 PWM circuit is same as shown in Figure 6.73. The output of the PWM circuit is given to a JK flip–flop IC that is connected to behave as a toggle flip–flop. The output of the flip–flop is passed through a dead time delay circuit as shown in Figure 6.73. Only here there are four sets of Schmitt gate segments, one for each signal line. The outputs from the dead time delay circuit is used to drive the primary and secondary-side switches of the phase-modulated PWM inverter.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Hardware bread-boarding

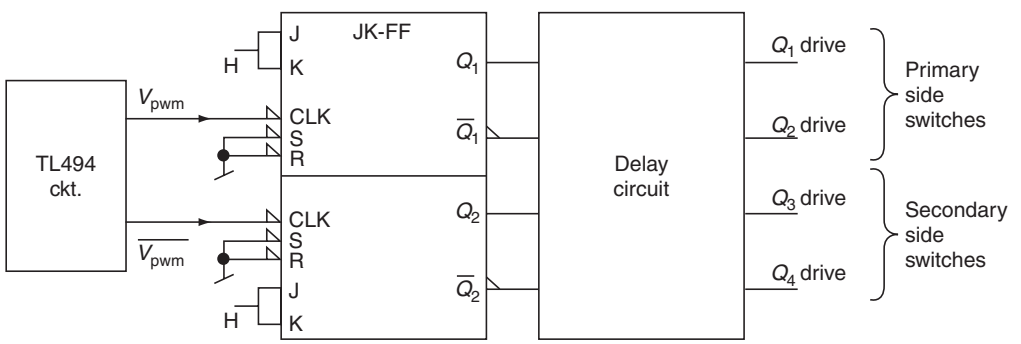


Figure 6.74 | Drive signal generation for phase-modulated PWM.

Tasks for study:

(a) Rig up the pulse-width modulator circuit followed by the toggle flip-flop and the dead time logic as shown in Figure 6.74.

(b) The PWM outputs, V_{pwm} and $\overline{V_{pwm}}$, are given as input to the JK flip–flop connected as a toggle flip–flop.

(c) Observe Q_1 and Q_2 drive signals on the oscilloscope both during the rising and falling transitions and note the dead time.

(d) Likewise observe the dead time between the Q_3 and Q_4 drive signals.

(e) Plot the waveforms of the TL494 outputs, flip-flop outputs and delay circuit outputs.

(f) Vary the modulating signal amplitude and frequency and observe the effect on the drive signals.

5. The circuit schematic shown in Figure 6.75 is a simple inverter for laboratory study wherein the top switches are *p*-channel MOSFETs and the bottom switches are *n*-channel MOSFETs. The DC link is a 15 V DC source. The *p*-channel MOSFETs can be driven with respect to the positive rail and the *n*-channel MOSFETs can be driven with respect to the negative rail. The load is a resistor connected across points A and B as shown.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice

b. Hardware bread-boarding

Tasks for study:

(a) Rig up the inverter circuit using *p*-channel and *n*-channel MOSFETs as shown in Figure 6.74.

(b) Using the PWM circuits of the previous laboratory experiments, modify the drive logic for the *p*-channel MOSFETs and apply the signals to the inverter switches.

(c) Observe the switching voltage waveform across Q_1 and Q_2 and that across Q_3 and Q_4.

(d) Observe the waveform during the switching transitions. If the *RC* time constant of the delay circuit is reduced, what is the effect during the switching transitions?

(e) Observe the voltage across the output load. If the load resistor is replaced by a resistor

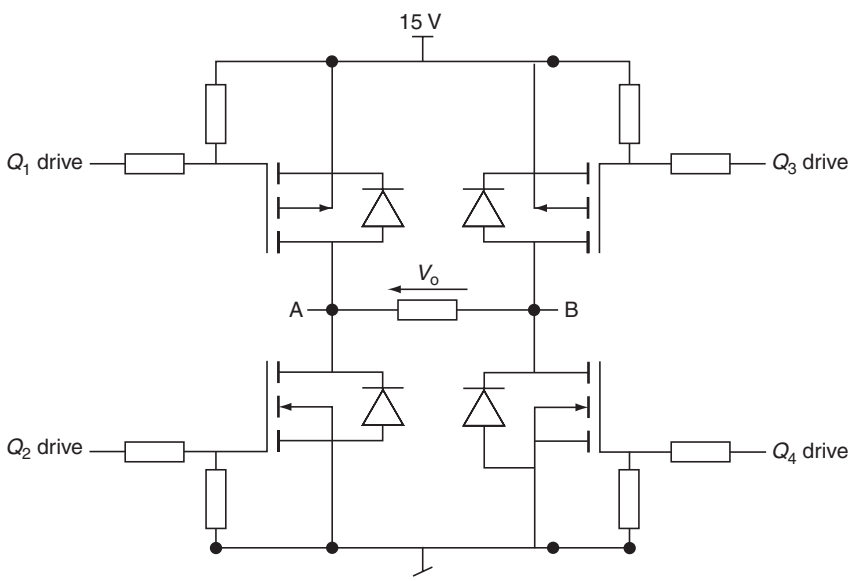


Figure 6.75 | Laboratory study inverter with simple non-isolated drive circuits.

in series with a capacitor, then what is the waveform across the capacitor? What is the effect of variation in the modulating signal amplitude and frequency on the output waveform?

(f) Apply bipolar and unipolar PWM to the above inverter and study the outputs with *R* load, *RC* load and *RL* load. What are the current freewheeling paths in the inverter for the *RL* load?

FILL IN THE BLANKS

1. DC–AC switched-mode converters are called _____.

2. The simplest inverter is composed of an _____ switch.

3. The three-phase inverter topologies consists of _____ bridge arms.

4. Switching transitions occur from the _____ of the core and the _____ characteristics of the power transistors.

5. In saturable-core inverter, the oscillation is not of the _____ type.

6. The ideal saturable-core inverter generates a frequency that is _____ proportional to the input DC voltage.

7. The saturable-core inverter generates a waveshape that is a _____ with a _____ duty cycle.

8. Better _____ can be achieved using the two transformer saturable base drive inverter as compared to the saturable output transformer topology.

9. In driven inverters, the power switches are driven through _____ control circuits where the ON-times and the OFF-times are _____.

10. The push–pull configuration is different from the other configurations in that the load is connected _____ of the SPDT switch.

11. The OFF-state voltage across the switches in the case of a push–pull inverter is _____ the input DC voltage.

12. The OFF-state voltage across the switches in a half-bridge inverter is _____ the input DC voltage.

13. The voltage applied across the bridge arms in a half-bridge inverter is _____ the input DC voltage.

14. The OFF-state voltage across the switches in a full-bridge inverter is _____ the input DC voltage.

15. The voltage applied across the bridge arms in a full-bridge inverter is _____ the input DC voltage.

16. For a given load power, the current rating required for the power devices in the half-bridge configuration is _____ that for the full-bridge configuration.

17. The quasi-square waveshape becomes a square waveform when the dead band angle is _____.

18. The efficiency of the duty cycle topology is _____ than that of the DC-link control topology.

19. The three-phase inverter topologies require _____ bridge structures.

20. The three-phase inverter topologies can be built using three _____ structures.

21. If each arm of a three-arm bridge inverter generates a _____ wave and is displaced with the waveform of the other two arms by _____, then a six-step waveform is obtained across the phases of the three-phase load.

22. The peak value of the six-step waveform that is obtained across the phases of the three-phase load is _____ times the input DC-link voltage applied to the inverter.

23. The current-controlled inverter uses the _____ to generate the switching signals for the switches in the inverter bridge arms.

24. One of the main drawbacks of the current-controlled inverter is the _____ in the switching frequency of the inverter switches.

25. In a current-controlled inverter as the hysteresis band is made smaller, the switching frequency becomes _____.

26. The main objective of the pulse-width modulation is to control the _____ component of the pulse-width modulated waveform with respect to the applied DC voltage.

27. The modulation control index gives a measure of the _____ component present in the output PWM waveform.

28. The modulation index gives a measure of the relative peak amplitude of the _____ signal with respect to the carrier signal.

29. If the positive part of the waveform is a mirror image of the negative part, then the waveform is said to be _____ about the time axis and such a waveform does not contain _____ harmonics.

30. If the average value of the waveform is zero, the Fourier series coefficient, _____ is also zero.

31. If a waveform has odd symmetry about the origin, the Fourier series coefficients for the _____ terms are all zero.

32. In the case of three-phase inverters, the PWM signal for an arm is generated by a reference modulating wave that is phase shifted _____ with respect to the reference modulating waves of the other two arms.

33. The ratio of the carrier frequency to the modulating frequency is called the _____.

34. Modulation index is defined as the ratio of the peak amplitude of the _____ signal to the peak of the _____ signal.

35. In sinusoidal PWM, for low values of frequency ratio, the relative harmonic content is _____.

36. To ensure that the number of pulses for the positive half and the negative half are same and symmetric the frequency ratio should be an _____ number.

37. If the frequency ratio for sinusoidal PWM is greater than 20, then the relative harmonic content of the bipolar PWM significantly _____.

38. In sinusoidal PWM, the harmonic content of the harmonic number equal to the frequency ratio is quite _____.

39. In a full-bridge inverter with sinusoidal PWM, the transformer across the bridge should be designed for the frequency of the _____ signal.

40. Space-vector modulation is a technique employed for _____ inverter systems.

41. Space-vector PWM provides a _____ DC-link voltage utilization compared to the sinusoidal PWM.

42. Space-vector PWM provides _____ relative harmonic content compared to the sinusoidal PWM.

43. The switching frequency of the inverter switches is _____ the carrier frequency in space-vector PWM.

44. The switching loss is _____ in space-vector PWM-based system compared to a sinusoidal PWM-based system with the same carrier frequency.

45. If a constant amplitude space vector rotates at a constant speed, then the resulting waveforms with respect to time on the three equally spaced spatial axes are _____ and displaced in time with respect to each other by _____.

46. If three-phase sinusoidal waveforms of frequency ω and displaced with respect to each other by 120° in time are applied at the three equally spaced spatial axes, then the resulting space vector is of _____ amplitude and rotates at a constant speed of _____.

47. In a quasi-square waveform, there is only the dead band angle that can be used as a parameter either to control the _____ amplitude or to eliminate a _____.

48. In harmonic elimination, at every notch, the voltage transits from either positive to negative or negative to positive resulting in _____.

49. Staircase PWM combines the principles of the _____ PWM and the harmonic elimination by _____ method.

50. Staircase PWM gives _____ modulation control index than the bipolar PWM.

DESCRIPTIVE QUESTIONS

1. What is the distinction between choppers, oscillators and inverters?

2. How are the inverters classified based on the manner in which the power semiconductor devices are switched?

3. Explain how a single pole double throw switch can be used as a two-level chopper and a two-level inverter.

4. How many SPDT switches are needed for a three- and four-level inverter? Explain.

5. What is the maximum number of the output voltage levels for a two-arm bridge inverter? If the inverter is supplied by only one input DC source, what is the possible number of levels for the two-arm bridge inverter?

6. Discuss the multi-level inverter structures.

7. Discuss the three-phase inverter topologies.

8. Distinguish between driven and self-driven inverters.

9. Explain the operation of the inverter wherein the output transformer is a saturable-core transformer.

10. How does the saturable-core inverter start?

11. What are the advantages of the saturable base drive inverter when compared to the saturable output inverter?

12. Explain the operation of the saturable base drive inverter.

13. What are the configurations generally used for driven inverter circuits?

14. Explain the operation of the push–pull inverter.

15. Discuss the current freewheeling through the body diodes of the switches in a push–pull inverter when the load is inductive.

16. Explain the operation of the half-bridge inverter.

17. Discuss the current freewheeling through the body diodes of the switches in a half-bridge inverter when the load is inductive.

18. Explain the operation of the full-bridge inverter.

19. Discuss the current freewheeling through the body diodes of the switches in a full-bridge inverter when the load is inductive.

20. Draw the quasi-square waveshape of amplitude 100 V and dead band angles of 60° and 30°.

21. Discuss the switching sequence for the switches of a full-bridge inverter to generate a quasi-square waveshape output.

22. Give a block schematic to generate the signals for the switches of a full-bridge inverter to generate a quasi-square waveform.

23. How is output regulation achieved for a quasi-square wave inverter? Explain the DC-link control and duty cycle control methods.

24. Discuss the three-arm bridge inverter used for three-phase loads.

25. Explain the gate drive signals that need to be provided for the switches of a three-phase inverter to generate a six-step waveform.

26. Explain the various modes of the six-step inverter operation.

27. Define modulation control index.

28. Define relative harmonic content.

29. Distinguish between modulation control index m and the modulation index M.

30. What are the various ways in which PWM is classified?

31. A waveform has a Fourier series given by

$$\frac{4V_{dc}}{\pi}\left(\sin\omega t+\frac{\sin 3\omega t}{3}+\frac{\sin 5\omega t}{5}+\frac{\sin 7\omega t}{7}+\cdots\right)$$

and another waveform has a Fourier series given by

$$\frac{4V_{dc}}{\pi}\left(\cos\omega t-\frac{\cos 3\omega t}{3}+\frac{\cos 5\omega t}{5}-\frac{\cos 7\omega t}{7}+\cdots\right)$$

Plot the two waveforms and explain the difference.

32. In a quasi-square waveform, if the dead band angle becomes zero, what happens to the waveform?

33. What is end pulse modulation?

34. What happens to the modulation control index when the angular end pulse width is 60° for an end pulse modulation? Why?

35. What is center pulse modulation?

36. What is the difference between unipolar and bipolar sinusoidal PWM waveform?

37. How can the size of the inverter's output transformer be reduced?

38. What is the principle of PWM through phase modulation? What are its advantages over sinusoidal PWM?

39. Discuss the generation of the drive signals for switches of the phase-modulated inverter?

40. What are the advantages of space-vector PWM?

41. What are the primary vectors in the case of a three-phase PWM system?

42. For a three-phase system, how many states can be defined and what are they?

43. What is space-vector hexagon? Into how many sectors can it be divided? Discuss.

44. Discuss the sequencing of the inverter states such that the inverter switching frequency is half the space-vector carrier frequency.

45. What is over-modulation?

46. In space-vector PWM, if the time spent in the zero state is split into two equal halves and appended on either side of the active bounding sector state times, then what is the effect on the harmonics of the modulating signal? What is the shape of the modulating signal? Discuss.

47. Discuss space-vector modulation for multilevel bridge arm structures for the inverter.

48. Discuss harmonic elimination method with multiple notches?

49. Discuss harmonic elimination method by staircase PWM?

50. What is the difference between harmonic elimination method with multiple notches and staircase PWM?

PROBLEMS

1. For a two-arm seven-level bridge inverter topology, what is the number of distinct source potentials required?

2. A saturable output transformer inverter is supplied from a 12 V battery. The output transformer core is a ferrite-based EE core with a core cross-sectional area of 31 mm^2. The total number of turns on the primary is 200. If the saturation flux density is 0.3 T, then calculate the switching frequency of the inverter.

3. A saturable output transformer inverter having an efficiency of 80% is driving a 25 W load. The energy is drawn from a 12 V battery. Find the current and voltage rating of the primary-side switches.

4. What is the modulation control index for a square waveform?

5. A quasi-square waveform has a dead band angle of 30°. What is its modulation control index?

6. For a quasi-square waveform having a dead band angle of 60°, calculate the modulation control index and tabulate the relative harmonic indices for the third, fifth, seventh and ninth harmonics.

7. For a quasi-square waveform having a dead band angle of 30°, tabulate the relative harmonic indices for the third, fifth, seventh and ninth harmonics. Compare with the relative harmonic indices of the previous problem and explain why the third and harmonics that are multiples of three are insignificant.

8. For an end pulse waveform having an angular end pulse width of 30°, calculate the modulation control index and tabulate the relative harmonic indices for the third, fifth, seventh and ninth harmonics.

9. For a center pulse waveform having an angular center pulse width of 30°, calculate the modulation control index and tabulate the relative harmonic indices for the third, fifth, seventh and ninth harmonics.

10. For a sinusoidal pulse-width modulation, the frequency of the triangular carrier is 20 kHz and the frequency of the modulating signal is 50 Hz. What is the slope of the triangular carrier?

11. For a sinusoidal PWM having a modulation index of 1 and the frequency ratio of 7, calculate and compare the modulation control indices for unipolar and bipolar PWM.

12. It is required to generate a space vector of amplitude 0.5 at an angle of 15° (anti-clockwise) from the *a*-axis. How much time should the inverter system spend in the bounding states to generate this equivalent space vector in a period of 50 μs? What are the bounding states?

13. It is required to generate a space vector of amplitude 0.5 at an angle of 150° (anti-clockwise) from the *a*-axis. How much time should the inverter system spend in the bounding states to generate this equivalent space vector in a period of 50 μs? What are the bounding states?

14. It is required to generate a space vector at an angle of 30° (anti-clockwise) from the *a*-axis such that the vector tip just touches the hexagonal boundary. How much time should the inverter system spend in the bounding states to generate this equivalent space vector in a period of 50 μs? What are the bounding states?

15. In a quasi-square waveform, if the fifth harmonic needs to be eliminated, then what is the dead band angle?

16. Determine the number of notches and the notch angles of a multiple notch harmonic elimination PWM pattern such that the 5th, 7th and 11th harmonics are eliminated.

17. Determine the number of levels and the heights of the levels of a staircase PWM pattern such that the 5th, 7th and 11th harmonics are eliminated.

ANSWERS

Fill in the Blanks

1. inverters
2. SPDT
3. three
4. magnetic characteristics; static
5. LC resonance
6. directly
7. symmetrical rectangular; 50
8. efficiency
9. external; controllable
10. across the throws
11. twice
12. same as that of
13. half
14. same as that of
15. same as that of
16. double
17. 0°
18. better
19. three-arm
20. two-level single-arm
21. square; 60°
22. 2/3
23. hysteresis comparator
24. variations
25. higher
26. fundamental
27. fundamental
28. modulating
29. symmetric; even
30. a_o
31. cosine
32. 120°
33. frequency ratio
34. fundamental; carrier
35. high
36. odd
37. low
38. high
39. modulating
40. poly-phase
41. higher
42. lower
43. half
44. lesser
45. sinusoidal; 120°
46. constant; ω
47. fundamental; specified harmonic
48. voltage reversal
49. sinusoidal; multiple notches
50. higher

CHAPTER 7

Design of Magnetics

Learning Objectives

After reading this chapter, you will be able to:

- understand the principles of magnetic components.
- design the magnetic components like transformers, inductors and current transformers (CTs) for a given application.

Magnetic components are an important and integral part of any power electronic system. The magnetic components are categorized into two broad classes: (a) energy-transfer devices and (b) energy-storage devices.

The *energy-transfer devices* transfer the power from one energy port called the primary to another energy port called the secondary without ideally storing or losing any energy in the process of transfer. These devices are called the *transformers*. The transformers are further classified into two sub-classes: (a) potential transformers (PTs) and (b) current transformers (CTs). In PTs, the energy or power that is drawn by the primary is from a voltage source. This is translated as a scaled voltage to the secondary load. In the case of CTs, the energy or power that is drawn by the primary is from a current source. This gets translated as a scaled current at the secondary.

The *energy-storage devices* store the kinetic energy by virtue of a current flowing through it. These devices are called *inductors*. This chapter will discuss both these classes of magnetic components and their design principles.

7.1 Magnetic Concepts

There are two fundamental laws of electromagnetism that are used to link the electric domain quantities and the magnetic domain quantities. They are:

1. Ampere's Law.
2. Faraday's Law.

These two laws are the governing principles based on which all other relationships in the magnetic domain are derived. Ampere's law links the current through the coil of the electric domain and the magneto-motive force (*mmf*) of the magnetic domain. A current flowing through a loop of conductor consisting of N turns produces a magnetic field of intensity H. This is depicted in Figure 7.1.

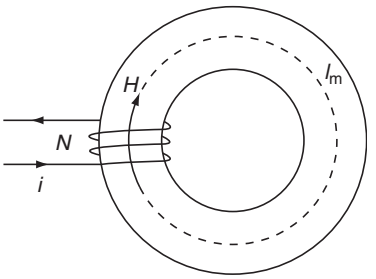


Figure 7.1 | Current through a loop producing magnetic field *H*.

The *Ampere's Law* is expressed as

$$\text{mmf} = \int_0^{l_m} H\mathrm{d}l = Ni \tag{7.1}$$

where H is the magnetic field intensity (A/m); l_m the magnetic path length (m); N the number of turns in the loop or coil; i the current in the coil. If the magnetic field intensity H is uniform along the magnetic path length, then Eq. (7.1) reduces to

$$Hl_m = Ni$$

$$H = \frac{Ni}{l_m} = \frac{mmf}{l_m} \tag{7.2}$$

Faraday's law links the voltage induced across the coil in the electric domain and the rate of change of flux in the magnetic core. Referring to Figure 7.1, any change in the flux ϕ in the core induces an electromotive force (emf) e across the coil. The induced voltage is proportional to the rate of change of flux in the magnetic core. *Faraday's law* is expressed as

$$emf = e = N\frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{7.3}$$

Figure 7.2 illustrates the energy interaction between the electric and the magnetic domains. The interaction between the two domains is through an energy port. Energy can flow from the electric domain to the magnetic domain and vice-versa. The energy or power variables in the electric domain are the voltage e and the current i. The corresponding power variables in the magnetic domain are the mmf and the rate of change of flux $\mathrm{d}\phi/\mathrm{d}t$.

From the Ampere's law given in Eq. (7.1),

$$N = \frac{mmf}{i}$$

Substituting for N in the Faraday's law given in Eq. (7.3), one obtains

$$e \times i = mmf \times \frac{\mathrm{d}\phi}{\mathrm{d}t} = \text{Power} \tag{7.4}$$

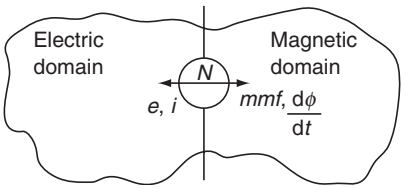


Figure 7.2 | Interaction between electric and magnetic domains.

Thus from Eq. (7.4), it can be observed that e is the potential variable, i is the flow variable in the electric domain, mmf is the potential variable and $d\phi/dt$ is the flow variable in the magnetic domain. Further observe that the product of the potential variable and the flow variable is always power. Thus, based on Eqs. (7.1), (7.3) and (7.4), the following analogies, as given in Table 7.1, can be expressed between the variables of the electric domain and that of the magnetic domain.

Dissipative Component

In Table 7.1, the resistance in the electrical domain is a dissipative component. It dissipates energy as heat. If i is the root mean square (rms) current flowing through R, then

$$i^2 R = \frac{1}{T}\int_0^T e \times i \times dt = \text{Power dissipated} \tag{7.4a}$$

Table 7.1 | Analogous variables in electric and magnetic domains

S. No.	Electric domain	Magnetic domain
1	e, v – voltage (volts)	mmf – magneto-motive force (amp-turns)
2	i – current (amps)	$d\phi/dt$ – rate of change of flux (weber/s)
3	Volts × Amperes = Watts	Ampere-turns × Weber/s = Watts
4	Resistance – $R = \dfrac{(1/T)\int v \times i \times dt}{i^2}$	$R_{mag} = \dfrac{(1/T)\int mmf\,(d\phi/dt)dt}{(d\phi/dt)^2} = \dfrac{(1/T)\int Hl_m d\phi}{(d\phi/dt)^2}$
5	Charge – $q = \int i dt$	$\phi = \int \dfrac{d\phi}{dt} dt$
6	Capacitance effect – $v = \dfrac{1}{C}\int i dt$	$Ni = mmf = \dfrac{1}{\Lambda}\int \dfrac{d\phi}{dt} dt = \dfrac{1}{\Lambda}\int d\phi$ where Λ is permeance $\Re = \dfrac{1}{\Lambda}$ = Reluctance

In the magnetic domain, the power dissipated is given as

$$P_{\text{dis}} = \frac{1}{T}\int_0^T mmf \times \frac{\mathrm{d}\phi}{\mathrm{d}t} \times \mathrm{d}t = \frac{1}{T}\int H \times l_{\text{m}} \times A_{\text{c}} \times \mathrm{d}B$$

where B_{m} is the maximum flux density discussed in the next sub-section. $\int H \times \mathrm{d}B$ is the area of the *B–H* loop. This means that in a period T the energy corresponding to the area of the *B–H* loop is dissipated. Then an equivalent dissipative component R_{mag} can be defined in the magnetic domain wherein

$$\left(\left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{\text{rms}}\right)^2 \times R_{\text{mag}} = \frac{V_{\text{c}}}{T}\int_{-B_{\text{m}}}^{B_{\text{m}}} H \times \mathrm{d}B \tag{7.4b}$$

where $V_{\text{c}} = A_{\text{c}} \times l_{\text{m}}$ is the core volume.

Flux and Flux Density

Figure 7.3 illustrates the concept of the flux density. The flow in the magnetic domain is the rate of change of flux. The magnetic core has a cross-sectional area A_{c} as shown. The flow $\mathrm{d}\phi/\mathrm{d}t$ is in a direction normal to the cross-sectional area. The flow per unit cross-sectional area is called the flow density. It is given as

$$\frac{1}{A_{\text{c}}} \times \frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\mathrm{d}(\phi / A_{\text{c}})}{\mathrm{d}t} = \frac{\mathrm{d}B}{\mathrm{d}t} \tag{7.5a}$$

where $B = \phi / A_{\text{c}}$ is called the flux density. The magnetic field intensity H gives rise to a certain flux density B (tesla) in the core. The relationship between B and H is given by

$$B = \mu H \tag{7.5b}$$

where μ is called the permeability of the core and is given as $\mu = \mu_{\text{o}} \cdot \mu_{\text{r}} \cdot \mu_{\text{o}}$ is the permeability of free space which is equal to $4\pi \times 10^{-7}$ H/m and μ_{r} is the relative permeability of the core material. Note that μ is not a constant quantity. It has been found to vary non-linearly with H. Graphically this relationship between B and H is as shown in Figure 7.4. Figure 7.4(a) shows the *B–H* curve that is linearized. The slope of the *B–H* curve is the permeability μ. Beyond a specific magnetic intensity, the flux density B clamps to a value B_{sat} and does not increase any further. At this instant the core is said to have saturated. When the core is in the saturated condition, the permeability μ is zero. When the core is saturated, for any change in the magnetic field intensity H, there is no change in B and therefore ϕ. Thus $\mathrm{d}\phi/\mathrm{d}t$ is zero and the power associated with the saturated core in the magnetic domain is zero. The hysteresis characteristics of practical cores are depicted

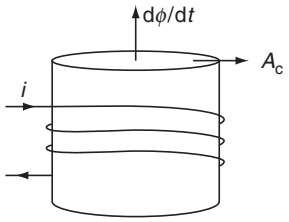


Figure 7.3 | Flux density.

Table 7.2 | Saturation flux density of common core materials

Core Type	B_{sat} (T)
Ferrite	0.3
Si steel	1.2 (CRGO), 1 (CRNGO)
Powdered iron	1.6
Amorphous glass	1.6
Mu metal	1

Notes: CRGO – cold rolled grain oriented; CNRGO – cold rolled non-grain oriented.

in Figure 7.4(b). Observe that a practical core exhibits *hysteresis*. The area within the hysteresis loop is the energy lost in every cycle and is called the hysteresis loss. The saturation flux density B_{sat} is a property of the material used. Table 7.2 gives the saturation flux density of common core materials.

The charge in the electrical domain is the integral of the flow variable, that is current. The integral of the flow variable in the magnetic domain is ϕ. From the Ampere's law [Eq. (7.2)] and the relationship as given in Eq. (7.5), one obtains

$$\mu \cdot A_c \cdot \phi = \frac{Ni}{l_m}$$

$$\phi = \frac{Ni}{l_m / (\mu_o \cdot \mu_r \cdot A_c)} = \frac{Ni}{\Re} = mmf \cdot \Lambda \tag{7.6}$$

where

$$\Re = l_m / (\mu_o \cdot \mu_r \cdot A_c) \Rightarrow \text{Reluctance}$$

$$\Lambda = \frac{1}{\Re} \Rightarrow \text{Permeance}$$

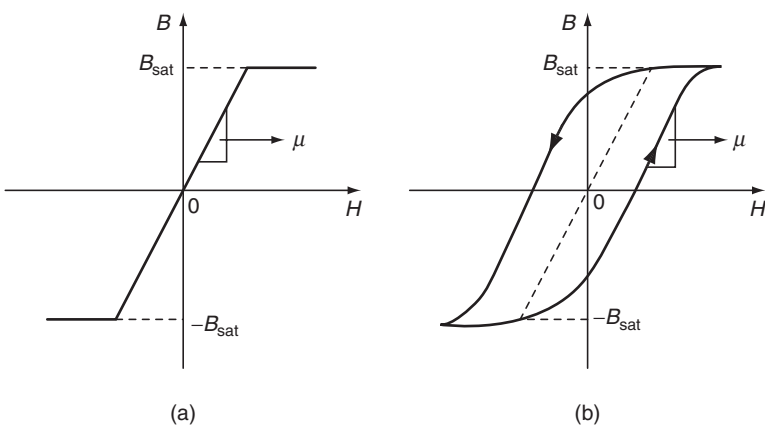


Figure 7.4 | (a) *B–H* curve (piece-wise linear); (b) practical *B–H* curve or *B–H* loop.

Volt-Second Balance

The flux ϕ can be expressed in terms of the properties of the core as

$$\phi = A_c B \tag{7.7}$$

The core cross-section area A_c is a constant for a specified core. However, the flux density varies depending on the level of the magnetic field intensity H. From the Faraday's law, the induced emf is given by Eq. (7.3). Using Eq. (7.7), the Faraday's law can be re-written as

$$e = NA_c \frac{dB}{dt} \tag{7.8}$$

The flux density B for the core can be obtained by integrating the above equation. Thus,

$$B = \frac{1}{NA_c} \int e \cdot dt \tag{7.9}$$

From Eq. (7.9) it can be observed that if the average value of e is non-zero, then $\int e \cdot dt$ accumulates towards infinity. This implies that the flux density B in the core continuously increases with time. However, from Figure 7.4 it is clear that there is a limit on the value of B, as given by B_{sat}. Therefore, for $\int e \cdot dt$ to be finite, the average value of e must be zero. If the frequency of the induced emf (e) is f, then the flux density that builds up in half a period can be estimated from Eq. (7.9) as

$$B = \frac{1}{NA_c}\left(\frac{2}{T}\int e \cdot dt\right)\frac{T}{2} = \frac{1}{NA_c} E_{avg} \frac{T}{2} \tag{7.10}$$

where $T = 1/f$; E_{avg} is the average voltage of e in a period $T/2$. Thus,

$$E_{avg} T = 2NA_c B \tag{7.11}$$

Equation (7.11) defines the volt-second product for a coil wound on a core. For a transformer of a given rated voltage, if a lower frequency is applied, the $E_{avg}T$ will increase and may saturate the core. When the core saturates, the windings offer zero impedance, which in turn causes huge currents to flow in the windings and the source, damaging either the source or the windings. For example, consider a transformer with the following rating: Primary: 230 V, 50 Hz.

If one applies 10 Hz at 230 V, then as the period at 10 Hz is higher than that at 50 Hz, from Eq. (7.11) it is clear that the core will tend to saturate. Therefore, one should never apply a voltage of frequency lower than that it is designed for. However, if the voltage is reduced proportionately, then a lower frequency can be applied. This means that, as given by Eq. (7.11), as long as one does not exceed the volt-time product or the *volt-second product* for the winding, the transformer will not saturate.

For one complete cycle the flux or the flux density should not build up. This means that the voltage across the coil should have a zero average over a full cycle. Thus

$$B = \frac{1}{NA_c}\left(\frac{1}{T}\int e \cdot dt\right)T; \quad \frac{1}{T}\int e \cdot dt = 0$$

The above equations should be satisfied to ensure that the flux density B does not build up over a full cycle. *This means that a magnetic winding will not support an average voltage across it. The average voltage across the magnetic winding under equilibrium conditions will always be zero. This is called the principle of volt-second balance.*

7.2 Inductor

Inductor Value

Equation (7.6) gives the relationship between flux, *mmf* and permeance. Thus,

$$\phi = \Lambda \cdot N \cdot i$$

Differentiating the equation and multiplying by N, one obtains

$$N\frac{\mathrm{d}\phi}{\mathrm{d}t} = \Lambda \cdot N^2 \cdot \frac{\mathrm{d}i}{\mathrm{d}t} \tag{7.12}$$

From the Faraday's law as given in Eq. (7.2), $N(\mathrm{d}\phi/\mathrm{d}t)$ is the induced voltage e across the coil. Thus,

$$e = \Lambda \cdot N^2 \cdot \frac{\mathrm{d}i}{\mathrm{d}t} = L \cdot \frac{\mathrm{d}i}{\mathrm{d}t} \tag{7.13}$$

Equation (7.13) is the derived form of Faraday's law as viewed from the electric domain. As the induced voltage e is the same whether viewed from the magnetic domain or the electric domain, the inductor value L is given by comparing Eqs. (7.12) and (7.13). L can be expressed in terms of the physical properties of the core by substituting the expression for permeance from Eq. (7.6) into Eq. (7.13):

$$L = \Lambda \cdot N^2 = \frac{\mu_\mathrm{o} \cdot \mu_\mathrm{r} \cdot A_\mathrm{c} \cdot N^2}{l_\mathrm{m}} \tag{7.14}$$

Energy Storage

In an inductor, the energy is stored by virtue of the current flowing through the inductor in the electrical domain. If L is the value of the inductor as viewed in the electrical domain, then the energy E_L stored in the inductor when a current i flows through it is given by

$$E_\mathrm{L} = \frac{1}{2}Li^2 \tag{7.15}$$

The same energy E_L as viewed within the magnetic domain is

$$E_\mathrm{L} = \frac{1}{2}Li^2 = \frac{1}{2}L\left(\frac{mmf}{N}\right)^2 = \frac{1}{2}\cdot\Lambda\cdot mmf^2 \tag{7.16}$$

where $\Lambda = L/N^2$ from Eq. (7.14). Thus, it can be observed from Eq. (7.16) that the energy-storage mechanism within the magnetic domain is a potential storage similar to that of capacitance storage in electric domain. Thus, the permeance is equivalent to a capacitance within the magnetic domain. From Eq. (7.6), the *mmf* dynamics is given as

$$mmf = \frac{1}{\Lambda}\cdot\phi = \frac{1}{\Lambda}\cdot\int \mathrm{d}\phi = \frac{1}{\Lambda}\cdot\int\frac{\mathrm{d}\phi}{\mathrm{d}t}\cdot\mathrm{d}t \tag{7.17}$$

Further, compare Eq. (7.17) with the following electric domain variables:

$$v = \frac{1}{C}\cdot q = \frac{1}{C}\cdot\int \mathrm{d}q = \frac{1}{C}\cdot\int\frac{\mathrm{d}q}{\mathrm{d}t}\cdot\mathrm{d}t = \frac{1}{C}\cdot\int i\cdot\mathrm{d}t \tag{7.18}$$

where q is the charge. From Eqs. (7.17) and (7.18), it can be observed that the permeance Λ is like a capacitance and the energy-storage mechanism *within the magnetic domain* is like a capacitive energy storage, that is, potential energy storage even though it appears as a kinetic or inductive energy storage as viewed in the electric domain. For a given core with a specified core cross-section area A_c and magnetic length l_m, the permeance is dependent on the permeability of the core μ. Beyond a particular value of magnetic field intensity, Ni/l_m, the core will saturate. The permeability of the saturated core is zero. Therefore, the permeance becomes zero and the core in saturated state is incapable of storing energy.

An air gap is generally introduced for inductive applications wherein the core needs to store energy. If an air gap of l_g is introduced in the magnetic path, then the reluctance of the magnetic path is given as

$$\Re = \frac{l_m}{\mu_o \mu_r A_c} + \frac{l_g}{\mu_o A_c} = \frac{1}{\mu_o A_c}\left(\frac{l_m}{\mu_r} + l_g\right) \tag{7.19}$$

The relative permeability μ_r is very large as compared to l_m. Therefore, $l_m / \mu_r \ll l_g$. The reluctance given in Eq. (7.19) can be written as

$$\Re \approx \frac{l_g}{\mu_o A_c} \tag{7.20}$$

The permeability μ_o is very small compared to l_g. This means that the reluctance will increase on introducing an air gap. Figure 7.5 depicts the change in the slope of the *B–H* curve on the introduction of the air gap within the core. The increase in reluctance reflects as a decrease in the effective permeability. Note that μ_g is the permeability of the core with the air gap of length l_g introduced. Referring to Figure 7.5, the core without air gap having a permeability of μ_1 will saturate at current i_1. However, on introducing an air gap the core will saturate at current i_g. Thus for a given energy-storage capability of the core, the core with air gap will handle higher *mmf* or, in other words, allow higher current capability for the windings.

Alternately, one may also argue that increasing the reluctance by introducing the air gap implies that the permeance has decreased. The energy-storing capability of the core is given by Eq. (7.16). Therefore, for a given energy-storage requirement, the permeance should decrease to handle higher *mmf*.

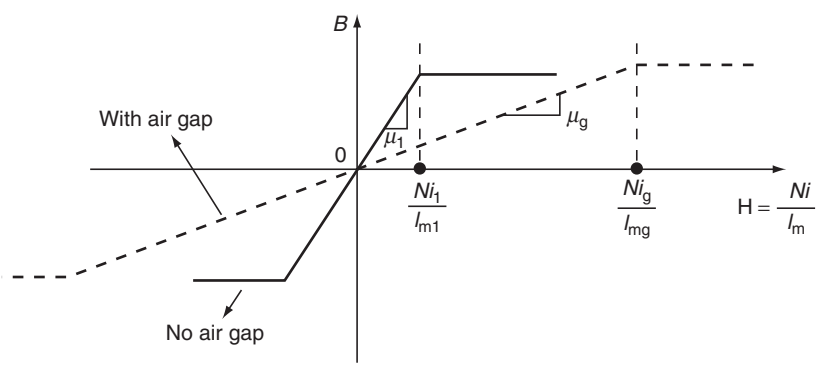


Figure 7.5 | Permeability change on introduction of air gap.

Area Product

An inductor consists physically of a winding and a magnetic core. The magnetic core includes the magnetic and physical properties of the material like the flux density, core cross-sectional area, window area, magnetic length, permeability, etc. On the other hand, the electrical side of the inductor defines the current through the inductor, the amount of energy that the inductor requires to store, the voltage across the inductor, etc. To design the inductor, the electrical and the magnetic parameters should be related to select the core, the winding gauge, the number of windings and the air-gap length that needs to be incorporated.

The area product approach is a systematic magnetic component design method. In this method, two physical parameters of the core, namely, the core cross-sectional area A_c and the window area A_w are related to the electrical parameters. Figure7.6 depicts the core cross-sectional area A_c and the window area A_w for a few standard core shapes.

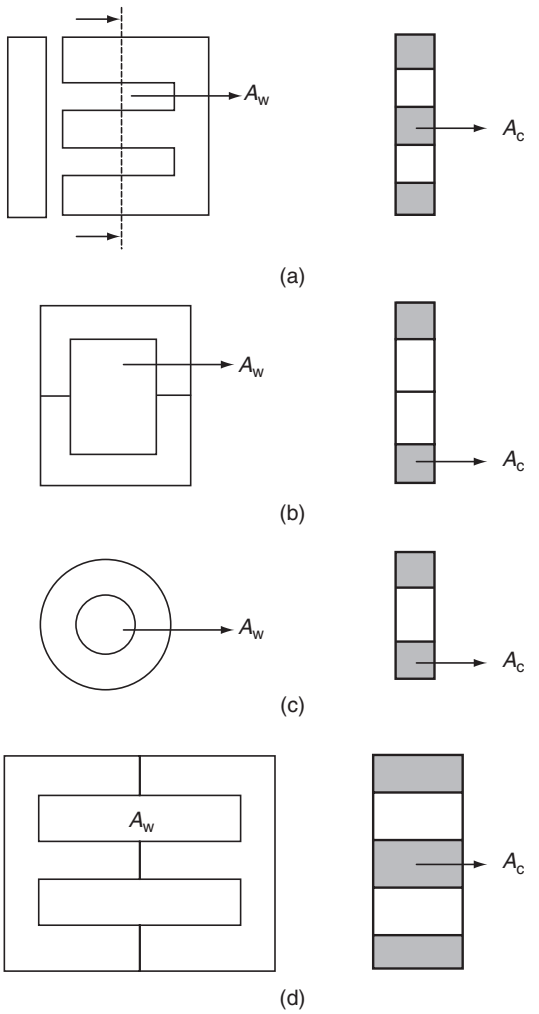


Figure 7.6 Illustration of the window area A_w and the core cross-sectional area A_c for a few core shapes: (a) EI core; (b) C core; (c) toroid; (d) EE core.

The Faraday's law of Eq. (7.8) relates the A_c to the voltage across the coil. The window area A_w is a measure of the number of turns of a given wire cross-section that can be accommodated. This cross-section is dependent on the current-carrying capacity or the rms current that flows through the wire. Thus A_w is related to the current capacity. The product of A_c and A_w is called the area product A_p. The area product should therefore be related to the energy that is stored in the core or the power that is transferred through the core. For energy-storing devices like the inductor, the area product A_p is related to the energy that the core will store. For energy-transferring devices like the transformers, the area product A_p is related to the power that will be transferred through the core.

Window Area

If "a" is the cross-section of the wire used and N is number of turns then the window area of the core should accommodate "N" turns of wire cross-section area "a". Thus

$$K_w A_w = N \cdot a = N \cdot \frac{I_{rms}}{J} \tag{7.22}$$

where K_w is the window utilization factor and has a value less than 1; J the current density (expressed as A/m^2).

The window utilization factor K_w accounts for the loss of the available window area due to various factors like insulation, air gaps, wire enamel, creepage, etc. The window utilization factor is discussed in detail in a later section.

The peak value I_m and the rms value I_{rms} of the current flowing through the inductor is related as

$$K_c = \frac{I_m}{I_{rms}} \tag{7.23}$$

where K_c is called the crest factor for the particular current waveshape through the inductor. Substituting Eq. (7.23) into Eq. (7.22), one obtains

$$K_w K_c A_w J = N \cdot I_m \tag{7.24}$$

Equation (7.24) relates the window area (A_w) parameter of the core to the electrical quantity which is the peak value of the current flowing through the inductor.

Cross-Section Area

From Faraday's law

$$e = L\frac{di}{dt} = N\frac{d\phi}{dt} = NA_c\frac{dB}{dt} \tag{7.25}$$

Integrating Eq. (7.25), one obtains

$$LI_m = NA_c B_m \tag{7.26}$$

where B_m is the maximum operating flux density in the core that corresponds to the maximum current I_m through the winding. Equation (7.26) relates the core cross-sectional area A_c to the current through the inductor and the flux density.

The energy that needs to be stored in the inductor is given as

$$E_{\mathrm{L}} = \frac{1}{2} L I_{\mathrm{m}}^{\ 2} \tag{7.27}$$

This can be re-arranged as

$$E_{\mathrm{L}} = \frac{1}{2} \cdot (L I_{\mathrm{m}}) \cdot I_{\mathrm{m}} \tag{7.28}$$

In the energy equation [Eq. (7.28)], LI_{m} is substituted with the core cross-sectional area relationship of Eq. (7.26) and I_{m} is substituted with the window area relationship of Eq (7.24). Thus,

$$A_{\mathrm{p}} = A_{\mathrm{w}} A_{\mathrm{c}} = \frac{2E_{\mathrm{L}}}{K_{\mathrm{w}} K_{\mathrm{c}} J B_{\mathrm{m}}} \tag{7.29}$$

Equation (7.29) is the basic area product equation for the energy-storing devices that is used for the selection of the core for a given energy-storage requirement and electrical/magnetic operating conditions.

Design of Inductors

The following steps illustrate the design for an energy-storing device, that is, the inductor.

Step 1 (Inductor Value): The value of the inductor has to be calculated. This is done based on the circuit topology where the inductor is used. For example, in the case of DC–DC converter applications, the inductor value is calculated based on the discussions in Chapter 5. For *LC* resonant circuits, the value of *L* is calculated based on the resonance frequency and the quality factor of the coil.

Step 2 (Area Product): From the current waveform through the inductor, the energy-storage requirement for the inductor is estimated. The energy that needs to be stored in the inductor is given as

$$E_{\mathrm{L}} = \frac{1}{2} L I_{\mathrm{m}}^{\ 2} \tag{7.30}$$

where E_{L} is the maximum energy in joules that needs to be stored in the inductor and I_{m} is the peak inductor current in amperes. Based on the energy to be handled by the inductor core, the area product is given as

$$A_{\mathrm{p}} = A_{\mathrm{w}} A_{\mathrm{c}} = \frac{2E_{\mathrm{L}}}{K_{\mathrm{w}} K_{\mathrm{c}} J B_{\mathrm{m}}} \tag{7.31}$$

where K_{W}, J and B_{m} are design parameters that have to be chosen by the designer for a particular application. The starting values or default values for these parameters are: Also, $K_{\mathrm{w}} = 0.6$ for single-winding inductors and 0.3 to 0.4 for multiple-winding inductors like the flyback transformer discussed in Chapter 5. Also $J = 3 \times 10^6$ A/m^2, $B_{\mathrm{m}} = 0.25$ T for ferrites, 1 T for cold rolled non-grain oriented (CRNGO) cores, 1.2 T for cold rolled grain oriented (CRGO) cores, 1.5 T for amorphous cores. The selected core should have an area product greater than that calculated by Eq. (7.31). Appendix II gives a representative list of commonly used core types and their physical parameters.

Step 3 (Permeance): The reluctance of a core with air gap l_{g} is given in Eq. (7.19). The permeance is the inverse of the reluctance and is given as

$$\Lambda = \frac{\mu_o \mu_r A_c}{l_m + \mu_r l_g} \tag{7.32}$$

where A_c is the core cross-sectional area of the selected core (in m^2); μ_r is the relative permeability of the selected core material; l_m is the magnetic path length of the selected core (in m); l_g is the introduced air-gap length (in m). The air-gap length is a design parameter that is chosen by the designer.

Step 4 (Number of Turns): From Eq. (7.14) it can be observed that the permeance is related to the inductance and the number of turns as

$$\Lambda = L / N^2$$

The permeance has the units of inductance per $turn^2$ or $H/turns^2$. Commercially, cores are available specifically for inductor applications wherein the gap is prefixed or the effective permeance is adjusted at the time of sintering the ferrite core material. The permeance of such cores is available in the data-sheet as A_L factor which is expressed in $nH/turns^2$. Re-arranging the above equation, one obtains

$$N = \sqrt{\frac{L}{\Lambda}} \tag{7.33}$$

Step 5 (Gauge of Wire): The cross-section area of the wire can be calculated as

$$a = \frac{I_{rms}}{J} \tag{7.34}$$

where I_{rms} is the rms value of the current flowing through the inductor which is equal to I_m/K_c. The gauge of the wire selected should have a cross-section area that is greater than that calculated by Eq. (7.34). Appendix III gives a representative list of commonly used wire gauges.

Step 6 (Available Window Area Check): The inequality $A_w K_w > aN$ should be satisfied or else repeat the calculations for the number of turns and gauge of wire after choosing the next bigger core. Note that the value of "*a*" used for checking the above inequality should be the actual cross-section area of the wire used and not the calculated value as per Eq. (7.34).

Problem 7.1

Design an inductor for a buck converter configuration as discussed in Chapter 5 for the following specifications:
Output voltage, $V_o = 3.3$ V
Output current, $I_o = 5$ A
Switching frequency, $f_s = 20$ kHz
Input voltage, $V_i = 10$ V ± 10%

Solution

Step 1: Determine *L*

The inductor value for this converter is given by Eq. (5.9) as

$$L = \frac{V_o \cdot (1 - D_{min})}{\Delta i_L \cdot f_s}$$

Δi_L is the current ripple in the inductor. Take Δi_L = 10% of I_o = 0.5 A.

As discussed in Chapter 5 and from Eq. (5.8), the minimum duty ratio is given as

$$D_{\min} = \frac{V_o}{V_{imax}} = 0.3$$

Substituting the appropriate values in the above equations, the value of the inductor is $L = 0.231$ mH

Step 2: Area Product

The energy and area product calculations are as follows

$$I_m = I_o + \frac{\Delta i_L}{2} = 5.25 \text{ A}$$

$$E_L = \frac{1}{2} L I_m^{\,2}$$

$$= \frac{1}{2} \times 0.231 \times 10^{-3} \times 5.25^2 = 3.1834 \times 10^{-3} \text{ J}$$

$$A_p = A_w A_c = \frac{2E_L}{K_w K_c J B_m}$$

As the switching frequency is 20 kHz, a core with high resistivity has to be chosen to reduce the eddy current losses. A ferrite core material would be an appropriate choice. Take $B_m = 0.25$ T for ferrite, $J = 3 \times 10^6$ A/m^2, $K_c \approx I_m / I_o$ and $K_w = 0.6$. Substituting the values in the area product equation, one obtains

$$A_p = 1.34747 \times 10^{-8} \text{ m}^4 = 13474.7 \text{ mm}^4$$

Select a core from Appendix II which has an A_p higher than the value calculated above. Let P 36/22 be selected ($A_c = 201$ mm^2, $A_w = 101$ mm^2, $A_p = 20100$ mm^4, $l_m = 53.2$ mm).

Step 3: Permeance

To calculate the permeance, the relative permeability is taken from the datasheet for the core material that is being used. A representative value from the datasheet for CEL HP_3C grade ferrite is given as

$$\mu_r = 2000 \pm 25\%$$

The relative permeability has a wide tolerance band. At the lower tolerance value of relative permeability, the permeance value is minimum which results in lower value of inductance for a specified number of turns. Therefore, for worst case design, use the lower tolerance value of relative permeability if the inductance value should be *atleast* the value as specified in Step (1). Use the higher tolerance value of relative permeability if the inductance value should be *atmost* the value as specified in Step (1). For this example the lower tolerance value of relative permeability is used for calculating the permeance.

The air-gap length is chosen as 0.5 mm. Incorporating these values in the permeance equation gives

$$\Lambda = \frac{\mu_o \mu_r A_c}{l_m + \mu_r l_g} = \frac{4\pi \times 10^{-7} \times 1500 \times 201 \times 10^{-6}}{53.2 \times 10^{-3} + 1500 \times 0.5 \times 10^{-3}} = 4.717 \times 10^{-7} \text{ H/turns}^2$$

Step 4: Number of Turns

The number of turns can now be calculated from the permeance and inductance relationship. It is given as

$$N = \sqrt{\frac{L}{\Lambda}} = \sqrt{\frac{0.231 \times 10^{-3}}{4.717 \times 10^{-7}}} = 22.13$$

Rounding up to the next higher integer if the calculation does not give an integer value, the number of turns N is 23.

Step 5: Wire Gauge Selection

The gauge of the wire is now estimated as follows: In Step (3) the value of current density J used is $J = 3 \times 10^6$ A/m^2. The cross-section area of the wire is calculated as

$$a = \frac{I_{\text{rms}}}{J} = \frac{5}{3 \times 10^6} = 1.67 \times 10^{-6} \text{ m}^2 = 1.67 \text{ mm}^2$$

Referring to Appendix III, select a wire gauge whose cross-section area is greater than that calculated above.

SWG 16 is selected ($a = 2.075$ mm^2)

Step 6: Available Window Area Check

The inequality $A_w K_w > aN$ has to be checked.

$A_w K_w = 101 \times 0.6 = 60.60$ mm^2 and $aN = 2.075 \times 23 = 47.725$ mm^2

The inequality is satisfied, which means that the windings will fit into the available window area.

Multiple-Winding Inductors

In certain applications, inductors can have more than one winding. One winding may be used to store energy into the core and another winding may be used to remove energy from the core. In some applications, the energy stored in the core may need to be shared with more than one load. In all such cases, the inductor will have multiple windings.

One of the more popular multiple-winding inductor is the flyback transformer. The flyback transformer is a misnomer. It is not a transformer in the sense that it is an energy-storing device by function. Therefore, the flyback transformer that is used in the flyback DC–DC converter topology (discussed in Chapter 5) is nothing but a multiple-winding inductor.

The design of the multiple-winding inductor is very similar to the single-winding inductor. The winding which is connected to the source that will transfer the energy to be stored in the core, is the main winding. The inductor design is done with respect to the main winding assuming that the main winding is the only winding. The difference is in deciding the number of turns for the other windings and the available window area as the window utilization factor is lower in multiple-winding inductors. The value of K_w is between 0.3 and 0.4 for multiple-winding inductors. All the design steps up to Step (3) are the same as for the single-winding inductor. There are minor modifications from Step (4) onwards. They are:

Step 4 (Number of Turns): Permeance is related to inductance and the number of turns as

$$N = \sqrt{\frac{L}{\Lambda}} \tag{7.35}$$

The number of turns N is for the main winding. The other windings called the subsidiary or the secondary windings are calculated based on the turns ratio obtained from the output voltage requirements and/or circuit topology. If $n_1, n_2, \ldots, n_m$ are the turns ratio of the m secondary windings that share the same energy of the core, then

$$N_{s1} = n_1 \cdot N, \ldots, N_{sm} = n_m \cdot N$$

Step 5 (Gauge of Wire): The cross-section area of the wire can be calculated as

$$a = \frac{I_{\text{rms}}}{J} \tag{7.36}$$

where I_{rms} is the rms value of the current flowing through the main winding of the inductor which is equal to $I_{\text{m}}/K_{\text{c}}$. If $I_{\text{rms-i}}$ is the rms value of the current flowing through the ith secondary winding, then

$$a_{s1} = \frac{I_{\text{rms-1}}}{J}, \ldots, a_{sm} = \frac{I_{\text{rms-}m}}{J}$$

Step 6 (Available Window Area Check): The inequality

$$A_{\text{w}} K_{\text{w}} > a \cdot N + \sum_{i=1}^{m} a_{si} \cdot N_{si} \tag{7.37}$$

should be satisfied or else repeat the calculations for the number of turns and gauge of wire after choosing the next bigger core. Note that the value of "a" and "a_{si}" used for checking the above inequality should be the actual cross-section areas of the wires used and not the calculated values as per Step (5).

7.3 Potential Transformer

Unlike an inductor, a transformer does not store energy. It only transfers the energy from one energy port to another. Therefore, a transformer must have at least two windings. The winding that is connected to the energy source is called the primary winding. The other winding(s) to which the load(s) is connected is called the secondary winding(s).

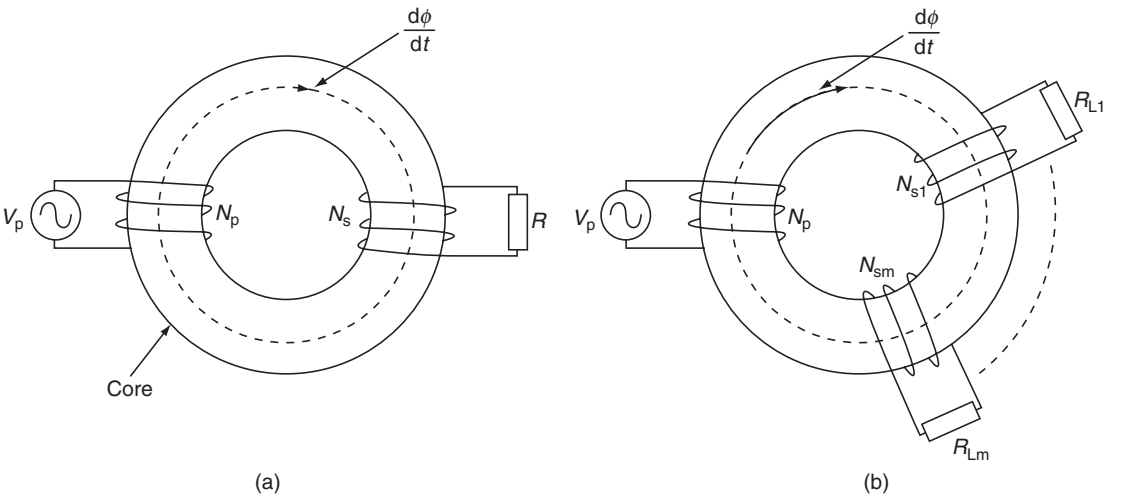


Figure 7.7 (a) Single secondary transformer; (b) multiple secondary transformer.

If the power source in the primary is a voltage source then the transformer is called a PT. If the power source in the primary is a current source then the transformer is called a CT. The principle of operation for both the transformers is the same though the causal variables are different. Figure 7.7(a) shows a basic transformer consisting of the primary winding N_p connected to a voltage source V_p, a secondary winding N_s connected to a load resistor R. Both the windings are wound on a common core that shares the same flux. Figure 7.7(b) shows a transformer with multiple secondary windings that are connected to individual loads. All the windings share the same core.

Operating Principle

On applying an AC voltage to the primary, current starts flowing through the primary winding. This produces a magnetic field around the coil, the intensity (H) of which is given by the Ampere's law. The magnetic field around the coil results in an *mmf* (Ni) within the core. The *mmf* causes a flux change within the core. All the windings linked to the core experience the flux change and by Faraday's law a voltage is induced across each secondary winding.

The operating principle of the transformer is illustrated in Figure 7.8. Consider an AC voltage source V_i applied to the primary winding N_p. A current i_p flows through the primary winding to set up an *mmf* in the core. The *mmf* due to the current flowing in the primary winding is given as

$$mmf_p = N_p \cdot i_p \tag{7.38}$$

The *mmf* in the core sets up a flux change rate $d\phi/dt$. The rate of change of flux is decided by the voltage across the winding that is connected to the voltage source. In this case, the primary winding is connected to a voltage source V_i. Therefore,

$$e_p = V_i$$

and from Faraday's law

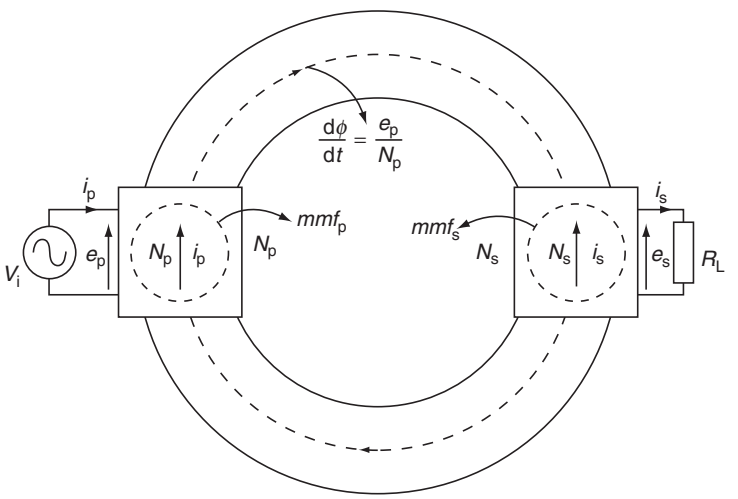


Figure 7.8 | Operating principle of the transformer.

$$\frac{d\phi}{dt} = \frac{e_p}{N_p} = \frac{V_i}{N_p} \tag{7.39a}$$

It can be observed from Eq. (7.39a) that the flux rate depends only on V_i and N_p and therefore is fixed by the input source voltage alone for a given number of primary turns. If a load resistor R_L is connected across the secondary winding, a current i_s flows through the load R_L. An opposing *mmf* is produced in the core by virtue of the secondary current flowing in the secondary winding. This *mmf* is given as

$$mmf_s = N_s \cdot i_s \tag{7.40}$$

If the load resistor R_L decreases, the secondary current i_s increases. This causes an increase in the opposing mmf_s. As $d\phi/dt$ is fixed by the input voltage source, the primary mmf_p will increase to maintain $d\phi/dt$ according to Eq. (7.39a) within the core. The mmf_p will increase by drawing more primary current i_p from the voltage source V_i. Thus more power is drawn from the source due to mmf_s acting as the loading effect on the primary mmf_p.

If there is no load on the secondary, then i_s is zero and therefore mmf_s is zero. The current drawn from the input source V_i is called the no-load primary current i_{po}. The current i_{po} drawn from the source now consists of two components: (a) i_m to set up the magnetic field; (b) i_c to supply the losses within the core like the hysteresis and eddy current losses. In most practical transformers the current drawn from the source under no-load condition is about 5–10% of the current at full load.

In the presence of the secondary load, the current drawn from the source consists of three components: (a) i_m to set up the magnetic field; (b) i_c to supply the core losses and (c) the load reflected current to account for the mmf_s.

The flux rate $d\phi/dt$ is caused *only* by the *mmf* developed due to the component of the primary current i_m that is used to set up the flux. This does not depend on the load as can be observed from Eq. (7.39a). Equation (7.39a) can be expressed in terms of circuit parameters like inductance and currents based on Eq. (7.13) as

$$\frac{di_m}{dt} = \frac{e_p}{L_M} = \frac{V_i}{L_M} \tag{7.39b}$$

where L_M is the magnetizing inductance and is given as $\Lambda \cdot N_p^2$ from Eqs. (7.12) and (7.13).

Figure 7.8 also shows the secondary winding that is wound over the same core. As the secondary winding is wound on the same core, the same flux rate also links the secondary windings. The voltage induced across the secondary winding according to the Faraday's law is given as

$$e_s = N_s \cdot \frac{d\phi}{dt} = N_s \cdot \frac{e_p}{N_p} \tag{7.41}$$

Turns Ratio

On re-arranging Eq. (7.41), one obtains

$$e_s = \frac{N_s}{N_p} \cdot e_p = n \cdot e_p \tag{7.42}$$

Equation (7.42) relates the primary and the secondary induced voltages. The variable "n" gives the ratio of the number turns in the secondary winding to the number of turns in the primary winding. Thus from Eq. (7.42),

$$n = \frac{N_s}{N_p} = \frac{e_s}{e_p} \tag{7.43}$$

Here "n" is called the *turns ratio*. The primary mmf_p has to account for the opposing load reflected mmf_s and the *mmf* required to set up the magnetic field and the core losses. Thus,

$$mmf_p = N_p \cdot i_p = N_s \cdot i_s + N_p \cdot i_{po} \tag{7.44a}$$

$$N_p \cdot \left(i_p - \frac{N_s}{N_p} \cdot i_s \right) = N_p \cdot i_{po}$$

$$i_p - n \cdot i_s = i_{po} \tag{7.44b}$$

Equation (7.44b) gives the no-load current that is used for setting up the magnetic field within the core and supplying the core losses.

Referring to Eq. (7.44a), if the transformer is designed such that $i_{po} << i_p$ then

$$N_p \cdot i_p \approx N_s \cdot i_s$$

and Eq. (7.43) can be extended to include the current ratio also as

$$n = \frac{N_s}{N_p} = \frac{e_s}{e_p} = \frac{i_p}{i_s} \tag{7.45}$$

Leakage Flux Linkage

The permeability of the core is somewhat similar to conductivity in an electric circuit. Just as no conductor is a perfect conductor of electricity, no magnetic material is a perfect conductor of magnetic field. As the permeability of a core is never infinity, but has a finite value, the permeance of the core and the air are like two parallel impedance paths in an electric circuit. One part of the flux rate flows within the core and another part of it flows in the air as indicated in Figure 7.9. The part of the flux rate that flows through the air path is called the leakage flux linkage as it does not link with any other winding except the primary winding.

For a current i_p flowing through the primary winding of N_p turns, the mmf_p that is set up is $N_p i_p$. This mmf_p causes a flux rate $d\phi/dt$ in the core of permeability $\mu_o \mu_r$ and a flux rate $d\phi_{\sigma p}/dt$ in the air path of permeability μ_o. From Faraday's law,

$$\frac{V_i}{N_p} = \frac{d\phi}{dt} + \frac{d\phi_{\sigma p}}{dt}$$

This can be re-arranged as

$$V_i = N_p \cdot \frac{d\phi}{dt} + N_p \cdot \frac{d\phi_{\sigma p}}{dt} \tag{7.46}$$

From Eqs. (7.13) and (7.44b), the above equation can be expressed equivalently with respect to circuit parameters like inductances and currents as

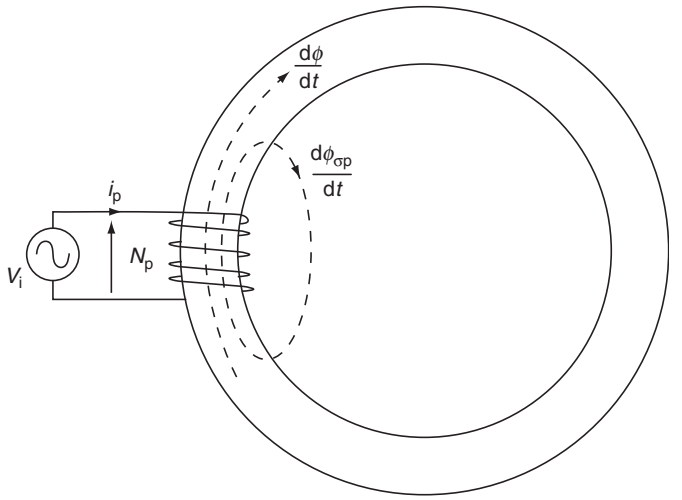


Figure 7.9 | Illustrating leakage flux linkage.

$$V_i = L_M \cdot \frac{di_m}{dt} + L_{\sigma p} \cdot \frac{di_p}{dt} \tag{7.47}$$

where the current i_m is responsible for setting up the field in the core, L_M is called the magnetizing inductance which represents the part of flux rate that flows through the core and $L_{\sigma p}$ is called the primary-side leakage inductance which represents the part of flux rate that flows through the air. The primary leakage inductance $L\sigma_p$ is given as $\Lambda_{\sigma p} \cdot N_p^2$ where $\Lambda_{\sigma p}$ is the permeance of the primary-side mean leakage path.

If the secondary winding is connected to a load, then a current i_s flows through it. The opposing mmf_s is $N_s i_s$. The mmf_s drives a flux rate $d\phi_{\sigma s}/dt$ to flow through the air which will not link with the primary winding. Thus, on the secondary winding,

$$\frac{e_s}{N_s} = \frac{d\phi}{dt} - \frac{d\phi_{\sigma s}}{dt}$$

This can be re-arranged as

$$e_s = N_s \cdot \frac{d\phi}{dt} - N_s \cdot \frac{d\phi_{\sigma s}}{dt} \tag{7.48}$$

From Eq. (7.13), the above equation can be expressed equivalently in terms of circuit parameters like inductances and currents as

$$e_s = \frac{N_s}{N_p} \cdot N_p \frac{d\phi}{dt} - N_s \cdot \frac{d\phi_{\sigma s}}{dt}$$

$$e_s = n \cdot L_M \cdot \frac{di_m}{dt} - L_{\sigma s} \cdot \frac{di_s}{dt} \tag{7.49}$$

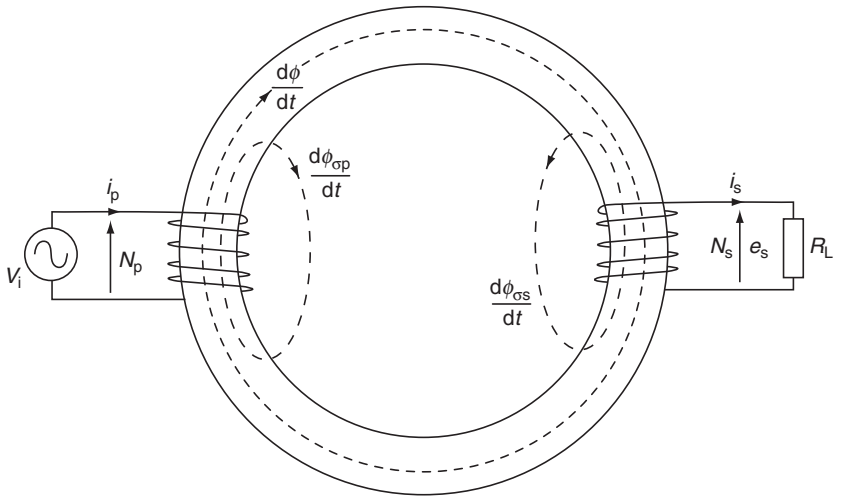


Figure 7.10 | Illustrating leakage flux linkage in primary and secondary.

where the current i_m is responsible for flux rate $d\phi/dt$ that links with the secondary windings, L_M is called the magnetizing inductance which represents the part of flux rate that flows through the core and $L_{\sigma s}$ is the secondary-side leakage inductance which represents the part of flux rate that flows through the air as depicted in Figure 7.10. The secondary leakage inductance $L_{\sigma s}$ is given as $\Lambda_{\sigma s} \cdot N_s^2$ where $\Lambda_{\sigma s}$ is the permeance of the secondary-side mean leakage path.

Equivalent Circuit

The transformer is a component wherein the energy flows from an electric domain through an energy port into the magnetic domain and from the magnetic domain to another electric domain through another energy port. It will be easier to visualize the operation of the transformer and obtain greater insights by replacing the transformer with an equivalent circuit representation. To develop the equivalent circuit, one can build equivalence for an ideal transformer to start with and proceed to extend the equivalence towards a practical transformer.

An ideal transformer has the following characteristics:

1. **Core with Infinite Permeance or Infinite Permeability:** This implies that the *B*–*H* curve is along the vertical axis. No magnetizing current is required to set up the magnetic field, that is, $i_m = 0$. Infinite permeance also implies that there is no leakage flux linkage. All flux linkages are through the infinite permeance core channel.
2. **Core with No Hysteresis Loop in the *B*–*H* Curve:** This implies that the hysteresis loss is zero.
3. **Core with Infinite Resistivity:** This implies that the eddy current loss is zero.
4. **Core which does not Saturate at any Magnetic Field Intensity:** This implies that the transformer can be connected to any voltage source of any frequency.
5. **Winding with Zero Resistance:** This implies that there are no winding losses and therefore the copper losses are zero.

The transformer equivalent circuit is developed starting with the ideal transformer. The equivalent circuit of the ideal transformer is shown in Figure 7.11(a). The variables e_p and e_s are the terminal voltages across the transformer primary and secondary windings, respectively, and i_p and i_s are the terminal currents. Here,

$$n = \frac{N_s}{N_p} = \frac{e_s}{e_p} = \frac{i_p}{i_s}$$

The magnetizing current $i_m = 0$ and the entire primary mmf_p is used to compensate the secondary mmf_s due to the secondary current flowing in the secondary winding and the external load resistance R_L. In an ideal transformer, the permeance is infinite. On introducing a finite permeance into the model, i_m is no longer zero. A finite i_m is required to set up the *mmf* in the core as the *B–H* loop has a finite slope. Figure 7.11(b) shows the equivalent circuit consisting of L_M followed by an ideal transformer of turns ratio n.
Referring to Figure 7.11(b),

$$V_i = e_p = L_M \frac{di_m}{dt}$$

where

$$i_m = i_p - n \cdot i_s$$

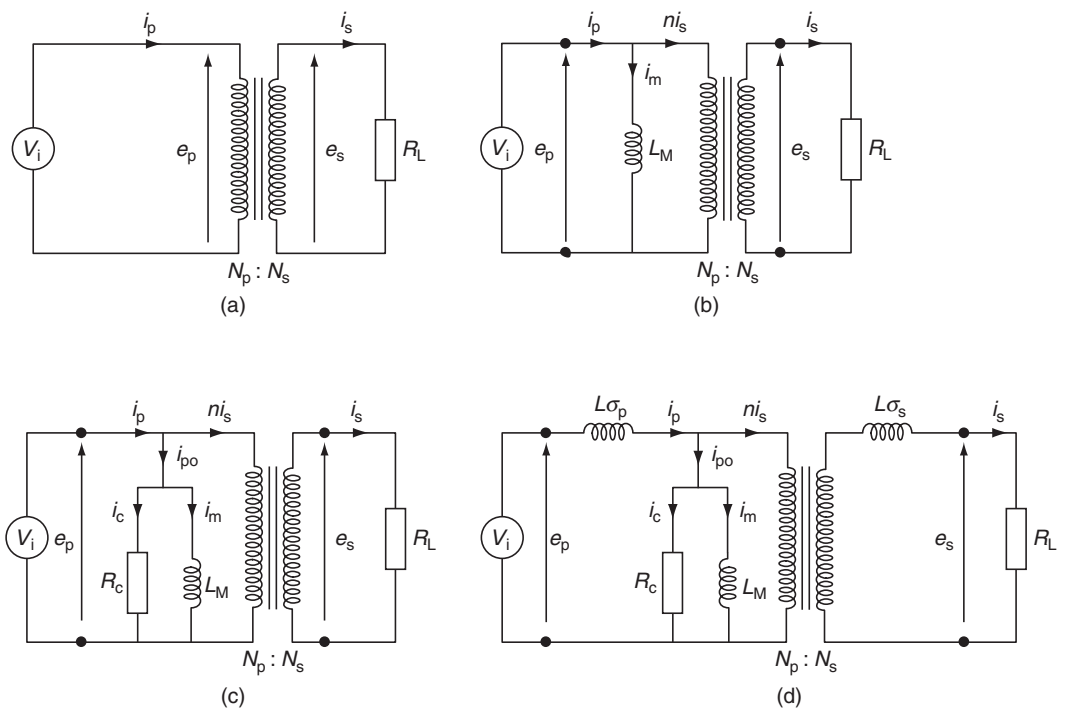


Figure 7.11 (a) Ideal transformer; (b) finite permeance non-ideality included – $i_m \neq 0$; (c) core loss non-ideality introduced; (d) leakage flux linkage non-ideality introduced.

Next the core loss non-ideality is introduced. This means that the core material exhibits *B–H* hysteresis loop characteristics giving rise to hysteresis loss. The resistivity of the core is also finite giving rise to eddy current loss. The core loss is depicted as a dissipative resistive element in the electrical domain as R_c in Figure 7.11(c). Here

$$i_p - n \cdot i_s = i_{po} = i_m + i_c$$

The no-load primary current i_{po} consists of the magnetizing component i_m and a core loss component i_c. Next the leakage flux linkage non-ideality is introduced. Once the permeance is assumed finite, leakage flux linkage path through the air gap will exist. This results in equivalent primary and secondary-side leakage inductances as shown in Figure 7.11(d). Here,

$$e_p = L_M \cdot \frac{di_m}{dt} + L_{\sigma p} \cdot \frac{di_p}{dt}$$

$$V_i = e_p$$

$$e_s = n \cdot L_M \cdot \frac{di_m}{dt} - L_{\sigma s} \cdot \frac{di_s}{dt}$$

$$e_s = i_s \cdot R_L$$

Finally, the winding resistance non-idealities are included. R_p is the primary winding resistance and R_s is the secondary winding resistance. The winding resistances give rise to the copper losses. The complete equivalent circuit of the transformer is shown in Figure 7.12. Here,

$$e_p = L_M \cdot \frac{di_m}{dt} + L_{\sigma p} \cdot \frac{di_p}{dt}$$

$$V_i = e_p + i_p \cdot R_p$$

$$e_s = n \cdot L_M \cdot \frac{di_m}{dt} - L_{\sigma s} \cdot \frac{di_s}{dt}$$

$$e_s = i_s \cdot R_L + i_s \cdot R_s$$

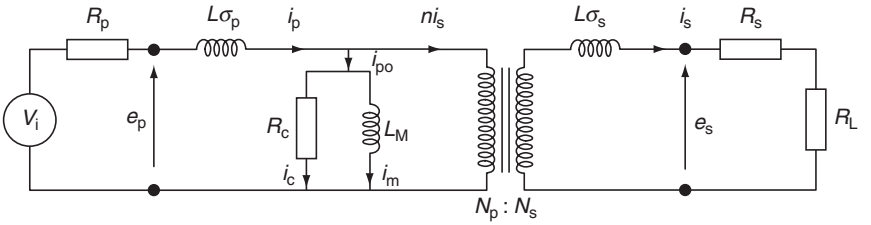


Figure 7.12 Complete equivalent circuit of the transformer with winding resistance non-idealities also introduced.

Area Product

A transformer consists physically of a primary winding and secondary windings on a common magnetic core. As discussed in the case of the inductor area product approach, here also the two physical parameters of the core, viz. the core cross-sectional area A_c and the window area A_w are related to the electrical parameters.

The Faraday's law of Eq. (7.8) relates the A_c to the voltage across the coil. The window area A_w is a measure of the number of turns of a given wire cross-section that can be accommodated. The wire cross-section is dependent on the current-carrying capacity or the rms current that flows through the wire. Thus A_w is related to the current capacity. The product of A_c and A_w is called the area product A_p. For energy-transferring devices like the transformers, the area product A_p is related to the power that will be transferred through the core.

Consider that the transformer has one primary winding of N_p turns and m secondary windings consisting of N_{s1}, N_{s2}, …, N_{sm} number of turns, respectively. The voltage across the primary winding is the source voltage V_i. From the Faraday's law,

$$e_p = N_p \frac{d\phi}{dt} = N_p A_c \frac{dB}{dt}$$

$$\frac{2}{T}\int_0^{T/2} e_p \cdot dt = \frac{2}{T} N_p A_c \Delta B \tag{7.50a}$$

where T is the period of the flux waveform and ΔB is the swing in the flux density during the half-period $T/2$.

$$\frac{2}{T}\int_0^{T/2} e_p \cdot dt$$

is the half-period average of the voltage across the primary winding. Thus, Eq. (7.50a) can be written as

$$E_{p\text{-}av} = 2 \cdot N_p \cdot A_c \cdot \Delta B \cdot f \tag{7.50b}$$

where $f = 1/T$ is the frequency of the flux waveform. From (7.50b), the number of turns in the primary winding is given as

$$N_p = \frac{E_{p\text{-}av}}{2 \cdot A_c \cdot \Delta B \cdot f} \tag{7.51}$$

where $E_{p\text{-}av}$ is the half-period average value of the voltage across the primary winding. Similarly, the number of turns in the secondary windings is

$$N_{si} = \frac{E_{si\text{-}av}}{2 \cdot A_c \cdot \Delta B \cdot f} \tag{7.52}$$

where N_{si} is the number of turns in the ith secondary winding; $E_{si\text{-}av}$ is the half-period average value of the induced voltage across the ith secondary winding.

The current flow through each winding causes copper losses, that is, power loss in the winding resistance. The winding resistance is a function of the wire length and the cross-sectional area of the wire. For a

given number of turns, the winding resistance is a function of the cross-sectional area of the wire. If the cross-sectional area of the wire is increased, then the winding resistance and hence the copper loss decreases. The cross-sectional area of the wire is determined by defining the current density J for the wire. Thus,

$$a = \frac{I_{\text{rms}}}{J} \tag{7.53}$$

where I_{rms} is the rms value of the current flowing through the winding and J is the current density for the winding in A/m^2.

For the primary winding, the wire cross-sectional area is given as

$$a_{\text{p}} = \frac{I_{\text{p-rms}}}{J} \tag{7.54}$$

and for the secondary windings, the wire cross-sectional areas are

$$a_{si} = \frac{I_{si\text{-rms}}}{J} \tag{7.55}$$

where $I_{si\text{-rms}}$ is the rms value of the current in the ith secondary winding and a_{si} is the wire cross-sectional area of the ith secondary winding. All the windings, that is the primary and the m secondary windings, should be accommodated within the core window area A_{w}. However, as discussed in the last section on inductors, the full window area A_{w} is not available as some area is needed to account for the wire enamel, coil former, insulation paper used between windings, etc. K_{w}, which has a value less than unity, denotes this loss of window area called the window utilization factor. Therefore,

$$K_{\text{w}} A_{\text{w}} = N_{\text{p}} a_{\text{p}} + \sum_{i=1}^{m} N_{si} a_{si} \tag{7.56}$$

Substituting for the wire cross-sectional area from Eqs. (7.54) and (7.55) into Eq. (7.56), one obtains

$$K_{\text{w}} A_{\text{w}} J = N_{\text{p}} I_{\text{p-rms}} + \sum_{i=1}^{m} N_{si} I_{si\text{-rms}} \tag{7.57}$$

Substituting for the primary and the secondary turns from Eqs. (7.51) and (7.52) into Eq. (7.57), one obtains

$$K_{\text{w}} A_{\text{w}} J = \frac{E_{\text{p-av}}}{2 \cdot A_{\text{c}} \cdot \Delta B \cdot f} I_{\text{p-rms}} + \sum_{i=1}^{m} \frac{E_{si\text{-av}}}{2 \cdot A_{\text{c}} \cdot \Delta B \cdot f} I_{si\text{-rms}} \tag{7.58}$$

$$A_{\text{p}} = A_{\text{w}} A_{\text{c}} = \frac{1}{2 \cdot K_{\text{w}} \cdot \Delta B \cdot f \cdot J} \left(E_{\text{p-av}} I_{\text{p-rms}} + \sum_{i=1}^{m} E_{si\text{-av}} I_{si\text{-rms}} \right) \tag{7.59}$$

If the form factor K_{f} is defined as the ratio of the rms value to the average value, then Eq. (7.59) can be written as

$$A_{\text{p}} = A_{\text{w}} A_{\text{c}} = \frac{1}{2 \cdot K_{\text{w}} \cdot K_{\text{f}} \cdot \Delta B \cdot f \cdot J} \left(E_{\text{p-rms}} I_{\text{p-rms}} + \sum_{i=1}^{m} E_{si\text{-rms}} I_{si\text{-rms}} \right) \tag{7.60}$$

where A_p is the area product in m^4. Equations (7.59) and (7.60) are the generic area product equations for transformers of any given source voltage waveshape. It should be remembered that the voltages are half-cycle average and rms values in Eqs. (7.59) and (7.60), respectively.

EXAMPLE 7.1

For a transformer that is supplied from a sinusoidal voltage source, the flux, voltage and current waveforms are sinusoidal in nature.

$K_f = 1.11$

The flux density swing in a half-period is from $-B_m$ to $+B_m$ where B_m is the maximum operating flux density and chosen to be less than B_{sat} for the core material. Therefore,

$\Delta B = 2B_m$

The sum of all the secondary VA is the consolidated load power P_o. Thus,

$$P_o = \sum_{i=1}^{m} E_{si\text{-rms}} I_{si\text{-rms}}$$

$$\frac{P_o}{\eta} = E_{p\text{-rms}} I_{p\text{-rms}}$$

where η is the efficiency of the transformer. Making the above substitutions into the generic area product Eq. (7.60), one obtains

$$A_p = \frac{P_o}{4.44 \cdot K_w \cdot B_m \cdot f \cdot J}\left(1 + \frac{1}{\eta}\right) \tag{7.61}$$

where P_o is the summation of all secondary VA; B_m the allowed maximum flux density in tesla; J the current density in A/m^2 which is commonly chosen as 3×10^6 A/m^2; K_w the window utilization factor which is chosen between 0.3 and 0.4 for practical transformers; f the frequency of operation in Hz.

Area Products for DC–DC Converters

Using the generic area product relationship given in Eq. (7.59), the area products of various DC–DC converters are worked out and tabulated in Table 7.3.

Table 7.3 | Area products of common DC–DC converters

Configuration	Area Product
Forward converter	$A_c A_w = \frac{P_o[1 + (1/\eta)]}{\sqrt{2} K_w J B_m f_s}$
Half-bridge converter	$A_c A_w = \frac{P_o[\sqrt{2} + (1/\eta)]}{4 K_w J B_m f_s}$

(*Continued*)

Table 7.3 | Continued

Configuration	Area Product
Full-bridge converter	$A_c A_w = \frac{P_o[\sqrt{2} + (1/\eta)]}{4K_w J B_m f_s}$
Push–Pull Converter	$A_c A_w = \frac{\sqrt{2} P_o[1 + (1/\eta)]}{4K_w J B_m f_s}$

Transformer Design

The previous section discussed the area product and the method to obtain the area product of the transformer for any given topology. This sub-section discusses the steps to design a transformer. The design steps are same whether the transformer is a low-frequency transformer or a high-frequency transformer.

Step 1 (Load Power Estimation): Before designing the transformer, it is necessary to estimate the total load power P_o that the transformer has to handle. This is calculated for a given topology as

$$P_o = \sum_{i=1}^{m} \left(\frac{1}{T} \int_0^T e_{si} \cdot i_{si} \cdot dt \right) \tag{7.62}$$

Equation (7.62) gives the summation of the load powers of all the m secondary windings.

Step 2 (Area Product Calculation): The area product for the transformer in a particular circuit configuration is estimated based on the generic area product given by Eq. (7.59) or Eq. (7.60). The area products for commonly used topologies can be worked out based on Eq. (7.59) and tabulated in a manner similar to that shown in Table 7.3 for DC–DC converters. Equation (7.61) gives the area product for sine wave transformers that are used in low-frequency grid applications. From the estimated value of the area product A_p, select a core from Appendix II that has an area product greater than that calculated.

Step 3 (Number of Turns): The number of turns is determined for the primary and the secondary windings as

$$N_p = \frac{E_{p\text{-}av}}{2 \cdot A_c \cdot \Delta B \cdot f} \tag{7.63}$$

(where $E_{p\text{-}av}$ is the half-period average value of the voltage across the primary winding)

$$N_{si} = \frac{E_{si\text{-}av}}{2 \cdot A_c \cdot \Delta B \cdot f} \tag{7.64}$$

(where N_{si} is the number of turns in the ith secondary winding and $E_{si\text{-}av}$ is the half-period average value of the induced voltage across the ith secondary winding.) The turns ratio for the ith secondary winding with respect to the primary winding is given as

$$n_i = \frac{N_{si}}{N_p} \tag{7.65}$$

Step 4 (Wire Gauge Selection): The wire gauges are selected based on the rms currents flowing in a given winding. The rms currents are first calculated based on the circuit topology on the primary and secondary sides. For the primary winding, the wire cross-sectional area is given as

$$a_{\mathrm{p}} = \frac{I_{\mathrm{p\text{-}rms}}}{J} \tag{7.66}$$

and for the secondary windings, the wire cross-sectional areas are

$$a_{si} = \frac{I_{si\text{-}\mathrm{rms}}}{J} \tag{7.67}$$

where $I_{\mathrm{si\text{-}rms}}$ is the rms value of the current in the ith secondary winding and a_{si} is the wire cross-sectional area of the ith secondary winding. From the values of the required estimate of the wire cross-sectional areas, one can select the wire gauge from Appendix III wherein the gauge selected is greater than that calculated.

Step 5 (Window Area Check): Wire gauges are available in discrete values as per the list given in Appendix III. One has to invariably select a wire gauge whose area is greater than the calculated one. This may necessitate higher window area requirement. When several windings are used (e.g., push–pull with primary and secondary center-tapped, multi-output SMPS, etc.), the window utilization factor K_{w} may decrease further as the various windings need more insulation layers. Using the *actual* selected wire cross-sectional areas, check to see if the turns fit into the effective window area $K_{\mathrm{w}}A_{\mathrm{w}}$ of the core by using the inequality:

$$K_{\mathrm{w}}A_{\mathrm{w}} > N_{\mathrm{p}}a_{\mathrm{p}} + \sum_{i=1}^{m} N_{si}a_{si} \tag{7.68}$$

If the above inequality is not satisfied, then select a core with the next higher A_{p} and re-do the calculations till the inequality is satisfied.

Problem 7.2

Design a transformer for the forward converter configuration with the following specifications:
Output voltage $V_o = 12$ *V*
Output current $I_o = 3$ *A*
Switching frequency $f = 20$ *kHz*
Supply voltage $V_i = 24\ \mathrm{V} \pm 10\%$
Refer to the discussion on forward converter in Chapter 5 to understand the operation of the converter.

Solution

Step 1 (Load Power Estimation): The secondary of the forward converter transformer has to supply the output load, the power dissipated in the secondary diodes and the power dissipated in the inductor winding and the inductor core losses. If the demagnetizing winding is used, then the power loss in the demagnetizing winding diode should also be accounted. However, as the current through the diode of the demagnetizing winding is the magnetizing current which is very small compared to the load current, the power loss in the demagnetizing winding can be neglected. Thus,

$$P_o = V_o I_o + V_D I_o + P_{L\text{-}cu} + P_{L\text{-}core} \tag{7.69}$$

where $P_{L\text{-}cu}$ is the copper loss of the secondary-side inductor winding and $P_{L\text{-}core}$ is the core loss of the inductor core. A starting conservative value of the inductor core and copper loss that is normally considered for the transformer design is $P_{L\text{-}cu} + P_{L\text{-}core} = 10\%$ of $V_o I_o$.

The converter diodes in the secondary of the transformer will show a significant drop as they are carrying high currents. The diode drops may be as high as 1.5 V for fast recovery diodes. It is safe to design for the worst case of $V_D = 1.5$ V. Equation (7.69) now becomes

$$P_o = 1.1 \times V_o I_o + V_D I_o = 1.1 \times 12 \times 3 + 1.5 \times 3 = 44.1 \text{ W}$$

Step 2 (Area Product Calculation): Referring to Table 7.3, the area product for the forward converter configuration is given by

$$A_p = \frac{P_o[1 + (1/\eta)]}{\sqrt{2} K_w J B_m f_s} \tag{7.70}$$

At high frequencies, usually the core material choice is ferrite. It has a saturation flux density B_s of 0.3 T. The maximum operating flux density in the core is chosen as 0.2 T ($B_m = 0.2$ T). Another important design parameter is the current density J. If it is chosen very low to bring down the copper loss, then for a given current, a very large conductor cross-section is required (thereby demanding a large window area). It should also be noted that a choice of very low current density will not have a significant reduction in the copper loss because at high frequencies the skin effect prevents decrease in the wire resistance. A current density between 2 and 5 A/mm^2 is found to be a good compromise between conductor resistance and window area. A value of $J = 3 \times 10^6$ A/m^2 is a default choice in most cases. The window utilization factor $K = 0.4$ and a conservative efficiency for the transformer is taken to be 0.8. Substituting the values in Eq. (7.70), one obtains

$$A_p = 1.46172 \times 10^{-8} \text{ m}^4 = 14617.2 \text{ mm}^4$$

Choose a suitable core from Appendix II that has an A_p greater than the value calculated above.

P 36/22 is selected ($A_c = 201$ mm^2, $A_w = 101$ mm^2, $A_p = 20100$ mm^4, $l_m = 53.2$ mm)

Step 3 (Number of Turns): The primary number of turns can be calculated using

$$N_p = \frac{E_{p\text{-}av}}{2 \cdot A_c \cdot \Delta B \cdot f}$$

For the forward converter, the half-period average value across the primary winding is V_{imax}, and ΔB is B_m. Thus

$$N_p = \frac{V_{imax}}{2 \cdot A_c \cdot B_m \cdot f} = \frac{26.4}{2 \times 201 \times 10^{-6} \times 0.2 \times 20000} = 16.4$$

Rounding to the next higher integer, $N_p = 17$ turns.

For the secondary, the secondary voltage and the output voltage are related by the forward converter input–output relationship as discussed in Chapter 5. Including even the non-idealities like the diode drop and the inductor winding resistance drop, one obtains

$$e_s \cdot D_{min} = V_o + V_D + 0.1V_o$$

$$V_{imin} = V_i - 10\%\ V_i = 21.6\ \text{V}$$
$$V_{imax} = V_i - 10\%\ V_i = 26.4\ \text{V}$$
$$D_{max} = 0.5$$
$$D_{min} = (V_{imin} \times D_{max})/V_{imax} = (21.6 \times 0.5)/26.4 = 0.41$$

$$e_s = \frac{V_o + V_D + 0.1V_o}{D_{min}} = \frac{12 + 1.5 + 1.2}{0.41} = 35.85\ \text{V}$$

$$N_s = \frac{e_s}{2 \cdot A_c \cdot B_m \cdot f} = \frac{35.85}{2 \times 201 \times 10^{-6} \times 0.2 \times 20000} = 22.3$$

Rounding to the next higher integer, $N_s = 23$ turns. The turns ratio is given as

$$n = \frac{N_s}{N_p} = \frac{23}{17} = 1.35$$

The demagnetizing winding N_d is equal to N_1 as they are wound bifilar.

Step 4 (Wire Gauge Selection): The rms values of the currents flowing in the windings are:

$$I_{s\text{-rms}} = I_o \sqrt{D_{max}} = 3 \times \sqrt{0.5} = 2.12\ \text{A}$$

$$I_{p\text{-rms}} = n \cdot I_{s\text{-rms}} = 1.35 \times 2.12 = 2.86\ \text{A}$$

The current through the demagnetizing winding is given as

$$I_{d\text{-rms}} = I_m \cdot \sqrt{\frac{1 - D_{max}}{3}}$$

where

$$I_m = \frac{V_{imin} \cdot D_{max}}{L_M \cdot f}$$

where

$$L_M = \frac{\mu_o \mu_r A_c N_p^2}{l_m}$$

A representative value from the datasheet for CEL HP_3C grade ferrite is given as $\mu_r = 2000 \pm 25\%$. The worst case I_m is obtained at minimum value of relative permeability. Therefore, taking $\mu_r = 1500$ and substituting in the L_M equation, one obtains

$$L_M = \frac{4\pi \times 10^{-7} \times 1500 \times 201 \times 10^{-6} \times 17^2}{53.2 \times 10^{-3}} = 2.06\ \text{mH}$$

The magnetizing current is given as

$$I_{\mathrm{m}} = \frac{21.6 \times 0.5}{2.06 \times 10^{-3} \times 20000} = 0.26 \text{ A}$$

$$I_{\text{d-rms}} = 0.26 \times \sqrt{\frac{1-0.5}{3}} = 0.106 \text{ A}$$

From the rms values of the currents and for a current density $J = 3 \times 10^6$ A/m^2, the required wire cross-sectional areas are

$$a_{\mathrm{p}} = \frac{I_{\text{p-rms}}}{J}, \; a_{\mathrm{s}} = \frac{I_{\text{s-rms}}}{J}, \; a_{\mathrm{d}} = \frac{I_{\text{d-rms}}}{J}$$

Substituting the values, one obtains
$a_{\mathrm{p}} = 0.953$ mm^2
Choose a wire gauge from Appendix III whose cross-section is greater than that calculated above: SWG 18 is a proper choice ($a_{\mathrm{p}} = 1.167$ mm^2).
$a_{\mathrm{s}} = 0.706$ mm^2
Choose a wire gauge from Appendix III whose cross-section is greater than that calculated above: SWG 19 is a proper choice ($a_{\mathrm{s}} = 0.8107$ mm^2).
$a_{\mathrm{d}} = 0.035$ mm^2
Choose a wire gauge from Appendix III whose cross-section is greater than that calculated above: SWG 35 is a proper choice ($a_{\mathrm{d}} = 0.03575$ mm^2).

Step 5 (Window Area Check): Now check for the inequality

$$K_{\mathrm{w}} A_{\mathrm{w}} > N_{\mathrm{p}} a_{\mathrm{p}} + \sum_{i=1}^{m} N_{si} a_{si}$$

Using the actual values of the cross-section areas, we have

$$N_{\mathrm{p}} a_{\mathrm{p}} + \sum_{i=1}^{m} N_{si} a_{si} = 17 \times 1.167 + 23 \times 0.8107 + 17 \times 0.03575 = 39.1 \text{ mm}$$

$A_{\mathrm{w}} K_{\mathrm{w}} = 40.4$ mm^2
Thus the above inequality is satisfied, which means that the windings will fit in the available window area.

7.4 Current Transformer

CTs are mainly used to sense the currents in the power circuits. The sensed currents are used as feedback signals for control, protection, monitoring and display. The CT is an energy-transferring device like the PT but the input energy source is a current source instead of a voltage source as is the case for the PT. However, all the magnetic principles that were discussed for the PT are applicable for the CT too.

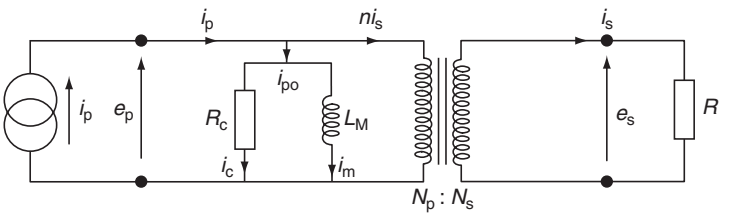


Figure 7.13 | Equivalent circuit of the CT.

Though the magnetic principles are same as that of the PT, the design constraints are different from that of the PT. This is due to the input energy source being a current source and the output being a voltage that is supposed to be proportional to the input current. Therefore the design methodology for the CTs is entirely different. CT is actually a dual of the PT. If the secondary load resistor R_L is made zero, then e_s is zero and therefore there is no flux flow rate within the core. If the secondary load resistor R_L were made infinity, that is, the secondary is open circuited, then there is no secondary current and as a result there is no secondary mmf_s ($= N_s i_s$). The only *mmf* in the core is the primary-side mmf_p given by $N_p i_p$. As there is no opposing *mmf*, the core may saturate.

The equivalent circuit for the CT is shown in Figure 7.13. It can be observed that the equivalent circuit is similar to that of the PT; however, the input energy source is a current source i_p. As the input energy source is a current source the series winding resistance and the leakage inductance terms are removed from the equivalent circuit as they will not affect the input current. Though these non-idealities are actually present, the design of CT will not consider these power losses and voltage drops as the main focus is on maintaining the current waveshape fidelity and not on efficiency.

In a CT, the primary number of turns is decided first. It is chosen as 1 or at most 2 in most cases irrespective of the size of the core. The secondary number of turns is decided by the circuit requirements and the maximum primary current value. N_s is much greater than N_p as the primary-side current is generally required to be stepped down.

If i_p is the primary current and i_s the required secondary current, then

$$n = \frac{N_s}{N_p}$$

and referring to the equivalent circuit as shown in Figure 7.13, one obtains

$$i_p - i_{po} = n \cdot i_s \tag{7.71}$$

$$i_s = \frac{i_p - i_{po}}{n} \tag{7.72}$$

If the secondary is terminated with a resistance R called the *burden* resistor, then

$$e_s = i_s \cdot R \tag{7.73}$$

The output voltage e_s across the sense resistor should be a measure of the source current i_p. However, from Eqs. (7.72) and (7.73) it can be observed that the output voltage would be an accurate measure of i_p if i_{po}

were equal to zero. Therefore in CTs it is imperative to ensure that i_{po} is very small compared to i_p. The current i_{po} consists of the magnetizing component i_m and the core loss component i_c. Reducing i_{po} implies that both the magnetizing component i_m and the core loss component i_c should be made as small as possible. If the rate of the flux or the magnetizing current is made small, then the absolute peak value of i_m will also be reduced. Therefore, from Faraday's law,

$$e_p = L_M \frac{di_m}{dt} \tag{7.74}$$

Rearranging the above equation,

$$\frac{di_m}{dt} = \frac{e_p}{L_M} = \frac{e_s}{n \cdot L_M} \tag{7.75}$$

In the case of switched-mode DC–DC converters, the currents are pulsed and so the secondary voltage developed by the pulsed secondary current is also pulsed. Therefore for the duration of the pulse, if the secondary voltage is assumed to be constant, then i_m is a linearly rising waveform. The maximum value of i_m occurs at the end of the current pulse width. If the current pulse width is t_w then integrating Eq. (7.75), one obtains

$$i_{m\text{-}max} = \frac{e_s \cdot t_w}{n \cdot L_M} \tag{7.76}$$

From Eqs. (7.75) and (7.76), it can be observed that the magnetizing current can be reduced by

1. decreasing the voltage across the secondary winding e_s;
2. increasing turns ratio n;
3. increasing magnetizing inductance L_M.

Ideally if e_s is zero then the magnetizing current $i_{m\text{-}max}$ is zero. However, if the CT secondary were short circuited then the secondary voltage would always be zero and as a result the primary objective of having a measure of the primary current for sensing purposes is not established. Therefore, e_s has to be finite but low (about 0.1–0.5 V). Also,

$$e_s = N_s \frac{d\phi}{dt} = \frac{N_s \cdot A_c \cdot \Delta B}{\Delta t} \tag{7.77}$$

Lower value of e_s implies that the flux density swing ΔB is also lower for a given core cross-sectional area and the number of turns in the secondary winding. This has the double effect of reducing both the magnetizing current i_m due to lower magnetic field intensity H requirement and also the core loss component as the B–H loop area is reduced. Thus for CT applications, B_m is chosen as around 10% of B_{sat} for the material. For a silicon steel CRGO core having a saturation flux density of 1.2 T, B_m is chosen as 0.12 T to keep the magnetizing component low.

Equations (7.75) and (7.76) also suggest that the magnetizing current can be reduced by increasing the turns ratio n and the magnetizing inductance L_M. The turns ratio is increased to the maximum by keeping $N_p = 1$. This will result in $n = N_s$. The magnetizing inductance is given as

$$L_M = \Lambda \cdot N_p^2 = \frac{\mu_o \mu_r A_c N_p^2}{l_m} \tag{7.78}$$

From Eq. (7.78), it can be observed that the magnetizing inductance can be increased by increasing the permeability of the core or the permeance of the core. Alternately, an oversized core with a larger core cross-sectional area can be used to increase the magnetizing inductance. Thus, one can conclude that for a good CT, the magnetizing current should be low and for this one needs

1. larger core cross-sectional area;
2. higher permeability core or core with larger permeance;
3. lower secondary voltage or lower burden resistance R;
4. large turns ratio n.

Figure 7.14 shows two schematics for processing the CT outputs. In Figure 7.14(a), the voltage across the burden resistor R is given as an input to a non-inverting operational amplifier (op-amp). In the case of Figure 7.14(b), the CT secondary dot polarity is reversed and connected to the op-amp as shown. The virtual ground principle of the op-amp is used in this circuit to make $e_s = 0$. The secondary current flows through the burden R that is connected across the output of the op-amp and the "–" terminal. In the case of CTs where the primary current is unidirectional, at the end of the duty cycle when the primary current becomes zero, the stored energy due to the magnetizing current will cause a large negative voltage spike at the secondary. To prevent such voltages reaching the load, a clamper circuit or a half-wave rectifier is used after the burden. In order to minimize the effects of the half-wave rectifier drop, a precision half-wave rectifier may be used.

Design of Current Transformers

The CTs are used mainly to measure or sense the current and are not used to transfer power. Therefore, in the case of CT design the issue is not one of power transfer but one of current waveshape fidelity. Therefore the design steps are different from those followed for the PTs.

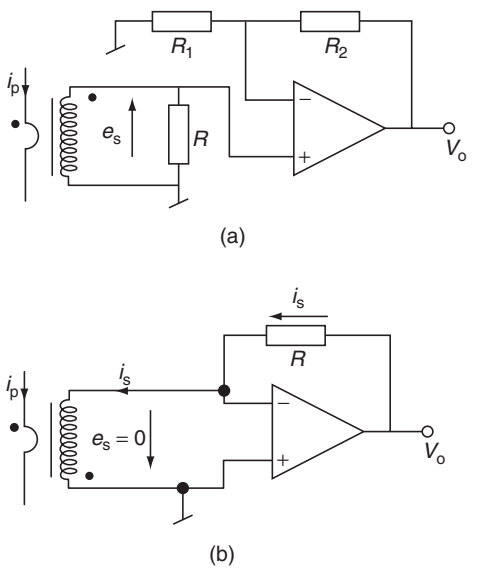


Figure 7.14 (a) CT output amplified by an op-amp non-inverting amplifier; (b) using virtual ground of op-amp to process the CT output.

Step 1 (Determine Turns Ratio): To start with choose $N_p = 1$ or at most 2. Based on the specification requirement decide on the required maximum secondary current, $i_{s\text{-max}}$ when the primary current is at the maximum value. Then,

$$n = \frac{i_{p\text{-max}}}{i_{s\text{-max}}}$$

The secondary number of turns is given as

$$N_s = n \cdot N_p$$

Step 2 (CT Burden): The secondary voltage e_s is a design specification. It should ideally be zero but should be chosen as a finite value. The value of e_s should be chosen as low as possible to obtain the best fidelity. The type of amplification and post-processing circuitry will determine the maximum value that can be chosen for e_s. As a thumb rule, choose $e_s < 0.1$ for very good fidelity and $e_s < 1$ for reasonable quality CTs. The output of the CT should be amplified by a non-inverting op-amp stage. The value of the burden resistor R is given as

$$R = \frac{e_{s\text{-max}}}{i_{s\text{-max}}}$$

Step 3 (Core Selection): In the case of CTs, the area product approach is not used as the objective here is not power transfer. Instead Eq. (7.77) is used which will estimate the core cross-sectional area based on the constraints of Steps (1) and (2). From the estimated core cross-sectional area an appropriate choice of core is made from Appendix II. Thus,

$$A_c = \frac{e_s \cdot \Delta t}{N_s \cdot \Delta B}$$

In the case of CTs where the input current i_p is bi-directional, the flux density swing in the core is from $-B_m$ to $+B_m$ in a maximum time of half the period. Therefore for CTs with bi-directional primary current,

$$A_c = \frac{e_s \cdot (T/2)}{2 \cdot N_s \cdot B_m} = \frac{e_s}{4 \cdot N_s \cdot B_m \cdot f}$$

In the case of CTs where the input current is uni-directional, the flux density swing in the core is from 0 to $+B_m$ in a maximum time of half the period. Therefore for CTs with uni-directional primary current,

$$A_c = \frac{e_s \cdot (T/2)}{N_s \cdot B_m} = \frac{e_s}{2 \cdot N_s \cdot B_m \cdot f}$$

Using the estimated value of the core cross-sectional area, select a core from Appendix II that has an A_c value greater than that calculated above.

Step 4 (Wire Gauge Calculation): The wire gauges are selected based on the rms currents flowing in a given winding. The rms currents are first calculated based on the current wave shape pattern. For the primary winding, the wire cross-sectional area is given as

$$a_p = \frac{I_{p\text{-rms}}}{J}$$

and for the secondary winding, the wire cross-sectional area is

$$a_s = \frac{I_{s\text{-rms}}}{J}$$

From the values of the required estimate of the wire cross-sectional areas, one can select the wire gauge from Appendix III wherein the gauge selected is greater than that calculated.

Step 5 (Window Area Check): In a manner similar to that discussed for the inductor and the PT design, one has to cross check whether the windings will fit into the available window area. Using the actual selected wire cross-sectional areas, check if the turns fit into the effective window area $K_w A_w$ of the core by using the inequality

$$K_w A_w > N_p a_p + N_s a_s$$

If the above inequality is not satisfied, then select a core with the next higher A_c and re-do the calculations till the inequality is satisfied.

7.5 General Notes on Magnetics

Window Utilization Factor

The window utilization factor K_w, has been used in both the inductor and transformer design to account for the non-availability of certain portion of the window area A_w. The effective window area that becomes available is $K_w A_w$ as discussed in the previous sections. The factors that affect the available window area are discussed in the following sub-sections.

Coil Former/Bobbin

These windings are generally wound over a bobbin or coil former. The bobbin that houses the windings has a certain thickness and this reduces the available window area. The unavailability of the window area due to the bobbin or coil former is accounted for by a factor K_{w1}.

Space Factor

There is an increase in the wire cross-sectional area due to a coating of insulating enamel on the bare copper wire. This results in a reduction in the winding space by a factor called the space factor K_{w2} defined as

$$K_{w2} = \frac{\text{Conductor area}}{\text{Conductor area} + \text{Insulation area}}$$

This factor depends on the wire gauge. For example, a SWG 45 wire gauge having a nominal diameter, d_{nom}, of 0.071 mm, and a maximum diameter with enamel insulation, d_{max}, of 0.086 mm.

$$K_{w2} = \frac{(\pi/4)d_{nom}^2}{(\pi/4)d_{max}^2} = \frac{0.071^2}{0.086^2} = 0.68$$

If the wire gauge is thicker, like say SWG 14, the value of K_{w2} comes out to be 0.91. Thinner the gauge, lesser is the space factor.

Air Gaps between Conductors

Another factor which reduces the available window area is that the adjacent turns are not air tight. There are always gaps between the conductors. This is due to the wires having a circular cross-section. In practice, this factor K_{w3} is found to be in the order of 0.6–0.9.

Insulation Factor

There is one more factor called the insulation factor, K_{w4}. Generally, when one winds a transformer, there are several insulation layers that are included, such as a layer between the primary winding and the secondary winding to meet the breakdown voltage requirements. If there are multiple secondary windings, additional layers have to be used further reducing the available window area.

Considering the above factors, the effective available window area is

$$\text{Available area} = K_w \times A_w$$

where

$$K_w = K_{w1} \times K_{w2} \times K_{w3} \times K_{w4}$$

Typically, a value of $K_w = 0.3$ to 0.4 can be assumed for transformer design purposes.

Core Shapes

There are numerous core shapes and sizes that are available commercially from many manufacturers. A few common core shapes will be illustrated here to give a flavor of the available commercial shapes.

Rod Core

This is normally made of ferrite and used in radios especially for tuning an inductor. The coil is wound on a cylindrical coil former which is inserted over the ferrite rod to form a solenoid as shown in Figure 7.15. The rod sits in the middle of the coil and adjustments of the rod position can be used to tune the inductance. The field spreads out into air at the ends of the rod. The path through the air ensures that the inductor does not saturate and remains linear. In this type of inductor radiation occurs at the end of the rod and electromagnetic interference (EMI) may occur.

"C" or "U" Core

Figure 7.16(a) shows a "U"-shaped core with sharp edges and Figure 7.16(b) shows a "C"-shaped core with rounded edges. Two UU or two CC cores can be coupled together to form a closed magnetic circuit. A U and I core or a C and I core can also be coupled to form a closed magnetic circuit.

"E" Core

The E-shaped cores are very popular. Some of the E core types are shown in Figure 7.17. Two EE cores of the same type can be coupled together to form a closed symmetrical magnetic circuit. Normally, the winding is wound around the center leg whose cross-sectional area is twice that of the individual outer legs. Figure 7.17(a) shows the classical E core that is most popular and commonly used. Figure 7.17(b) illustrates

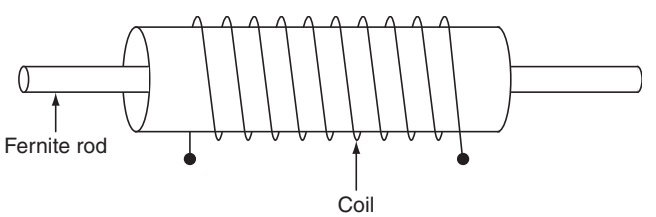


Figure 7.15 | Rod core.

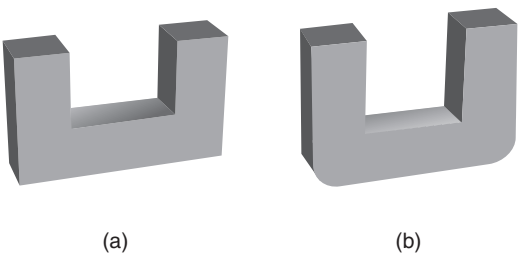


Figure 7.16 | (a) U-shaped core; (b) C-shaped core.

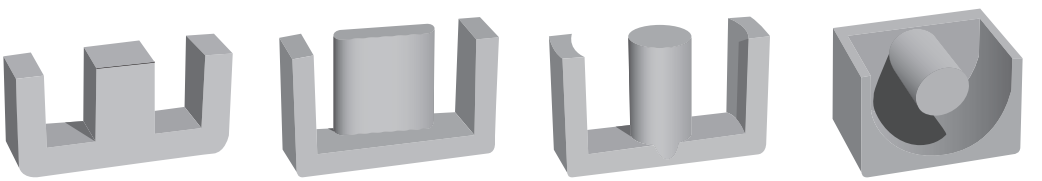


Figure 7.17 | (a) Classical E core; (b) EFD core; (c) ER core; (d) EP core.

the EFD core that is used for flat or low-profile inductors or transformers. Figure 7.17(c) shows the ER core that has a cylindrical central leg. Figure 7.17(d) shows and EP core that is in between an ER core and a pot core.

Two "E" cores will form a three-legged structure when coupled together as shown in Figure 7.18. If an air gap is required, the center leg of the "E" is shortened so that the air gap sits in the middle of the coil to minimize fringing and reduce EMI.

"E" and "I" Laminations

Sheets of silicon steel are stamped out in shapes like "E" and "I". These E and I stampings are stacked to form three-legged laminated core structures. The coils are wound on these three legs to form single-phase low-frequency transformers or low-frequency inductors or three-phase low-frequency transformers.

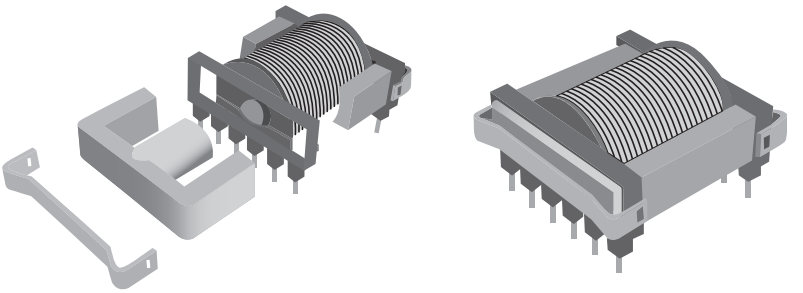


Figure 7.18 | Coupling two E cores.

Pot Core

This is used for inductors and transformers. The shape of a pot core is round with an internal hollow that almost completely encloses the coil former and coil. The pot core is made in two halves which fit together around the coil former (bobbin). This shape of core has a shielding effect, preventing radiation and reducing EMI. Figure 7.19 shows one half of a pot core.

Toroidal Core

The toroidal core is shown in Figure 7.20. The coil is wound through the hole in the toroid and around the outside. The coil is distributed uniformly all around the circumference of the toroid. The circular core geometry will naturally keep the field constrained within the core material. It is a low-radiation transformer and popular in amplifiers where the desirable features are high power, small volume and minimal EMI. It is also used in CTs for current-sensing applications. However, it is not easy to wind a coil on a toroidal core.

Planar Core

A planar core consists of two flat pieces of magnetic material as shown in Figure 7.21. One-half is placed above and another half is placed below a winding that is printed on a printed circuit board. It is typically used for PCB mountable applications. This design is excellent for small volume, low-profile and very high frequency transformers.

Polarity and Dot Convention

Knowledge of polarities and dot convention is very important when working with transformers. While making transformer connections, if the winding polarities are not carefully considered, then the results can be quite disastrous. Therefore the dot convention to mark the winding polarity must be strictly practiced both in symbolic circuit diagrams and also while making interconnections.

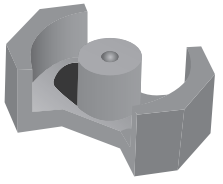


Figure 7.19 | Pot core.

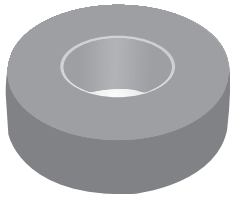


Figure 7.20 | Torodial core.

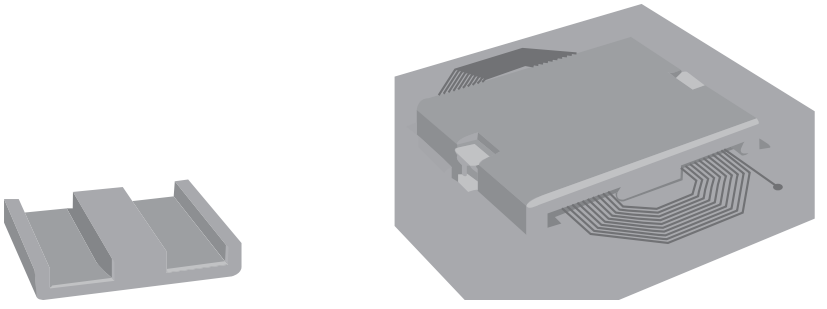


Figure 7.21 | Planar core and its mounting on a PCB.

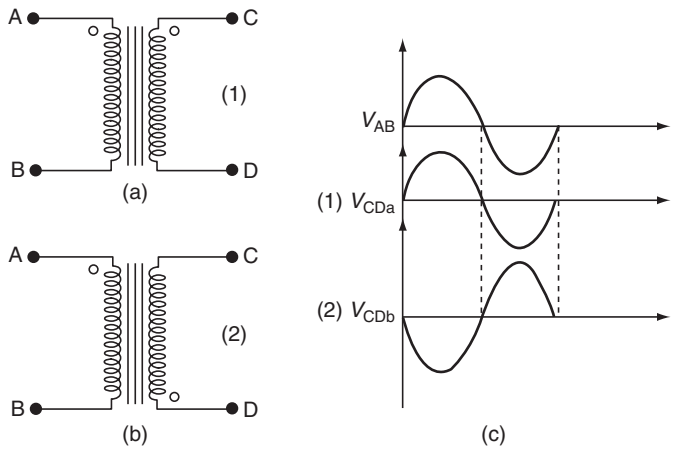


Figure 7.22 (a) Transformer with A and C having same polarity; (b) transformer with A and D having same polarity; (c) the waveforms for configurations shown in (a) and (b).

If two coils AB and CD are wound over a core in the same direction (clockwise or anti-clockwise), then points A and C will have the same phase with respect to the other end of the winding. In other words, if at any instant, point A is positive with respect to point B, then point C is also positive with respect to point D and vice-versa. Points A and C are said to have the same polarity. Symbolically, the same polarity ends of the windings of a core are denoted by a "dot" mark at the beginning of the winding as shown in Figure 7.22(a). Figures 7.22(a) and 7.22(b) show two transformers with different dot polarities. Figure 7.22(c) shows the winding waveforms for the two transformers.

In a transformer, when polarities are not known, it is possible to find out the polarities with a simple test. The transformer can be connected as shown in Figure 7.23(a). Connect any two ends, say BD, together and measure the input voltage V_1 and the voltage between A and C, that is, V_2. If the measured V_2 is less than V_{AB} ($V_2 = V_{AB} - V_{CD}$), then mark polarities as shown in Figure 7.23(b). If the measured V_2 is greater than V_{AB} ($V_2 = V_{AB} + V_{CD}$), then mark the polarities as in Figure 7.23(c).

Magnetic Losses

Hysteresis Loss

Referring to Figure 7.24, as one travels along the magnetization path of a virgin core specimen from O to B, the flux density swings from 0 to B_{sat}. If the magnetic field intensity H is again brought back to O, there exists some finite flux density (point C) or magnetic charge in the core due to the *magnetic retentivity* property of the core material. A negative magnetization field has to be applied to bring the specimen back to zero flux density (point D).

From Ampere's law, as given by Eq. (7.2), the magnetizing field H is

$$H = \frac{Ni}{l_m} = \frac{mmf}{l_m}$$

and the flux density is given by

$$B = \mu H$$

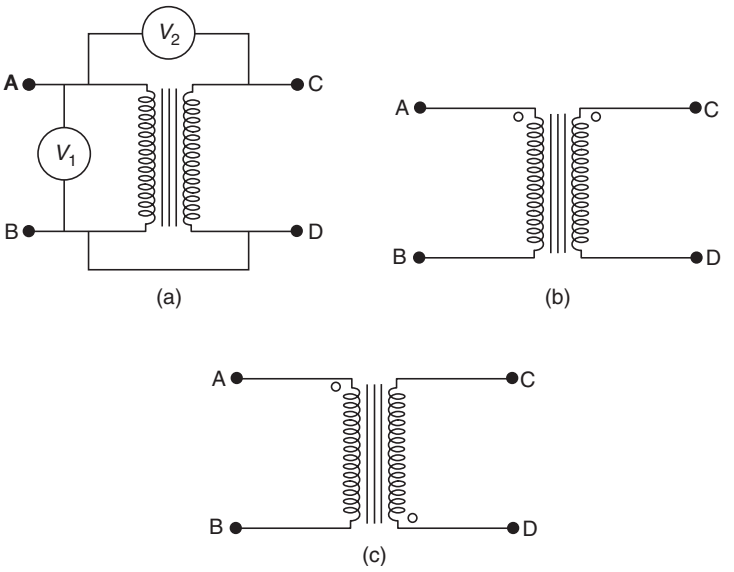


Figure 7.23 | Testing for dot polarity.

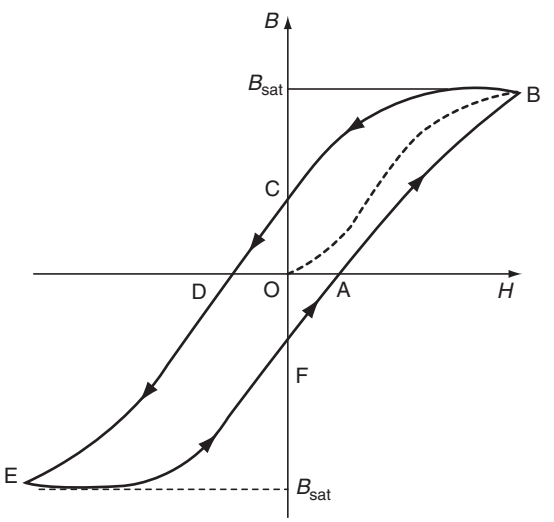


Figure 7.24 | *B–H* loop for a core material showing the hysteresis loop.

The induced voltage across a winding that is wound on the core is

$$e = N\frac{\mathrm{d}\phi}{\mathrm{d}t} = NA_{\mathrm{c}}\frac{\mathrm{d}B}{\mathrm{d}t}$$

From the above, the energy required to traverse the *B–H* loop in the path EFAB is

$$\begin{aligned} E_{\text{up}} &= \int_{0}^{T/2} mmf \cdot \frac{\mathrm{d}\phi}{\mathrm{d}t} \cdot \mathrm{d}t \\ &= \int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot l_{\text{m}} \cdot A_{\text{c}} \cdot \mathrm{d}B = V_{\text{c}} \int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot \mathrm{d}B \end{aligned} \tag{7.71}$$

where $V_{\text{c}} = l_{\text{m}} \cdot A_{\text{c}}$ is the core volume. Figure 7.25(a) illustrates the area of the *B–H* loop that represents the energy E_{up} where $\int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot \mathrm{d}B$ is the area under the *B–H* loop with reference to the *B*-axis.
Similarly, the energy required to traverse the down-swing of the *B–H* loop is given as

$$\begin{aligned} E_{\text{down}} &= \int_{T/2}^{T} mmf \cdot \frac{\mathrm{d}\phi}{\mathrm{d}t} \cdot \mathrm{d}t \\ &= \int_{B_{\text{m}}}^{-B_{\text{m}}} H \cdot l_{\text{m}} \cdot A_{\text{c}} \cdot \mathrm{d}B = -\int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot l_{\text{m}} \cdot A_{\text{c}} \cdot \mathrm{d}B = -V_{\text{c}} \int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot \mathrm{d}B \end{aligned} \tag{7.72}$$

Figure 7.25(b) illustrates the energy per unit volume that represents E_{down}. The total energy for one complete traversal of the *B–H* loop is given as

$$E_{\text{hys}} = E_{\text{up}} + E_{\text{down}} = l_{\text{m}} \cdot A_{\text{c}} \left(\underbrace{\int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot \mathrm{d}B}_{\text{Path EFAB}} - \underbrace{\int_{-B_{\text{m}}}^{B_{\text{m}}} H \cdot \mathrm{d}B}_{\text{Path BCDE}} \right) \tag{7.73}$$

Equation (7.73) gives the amount of energy required to traverse the *B–H* loop. It can be observed from Figure 7.25 that the energy required to traverse the *B–H* loop is equal to the product of the area of the *B–H*

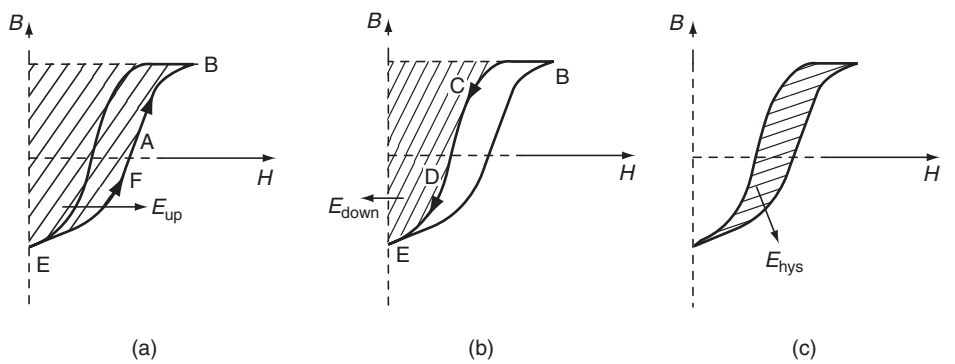


Figure 7.25 | (a) E_{up} Per unit volume; (b) E_{down} per unit volume; (c) E_{hys} per unit volume.

loop and the core volume $l_m \cdot A_c$. This energy is lost forever and is called the hysteresis loss. The power lost is the rate of energy required to traverse the *B–H* loop and is given as

$$P_{hys} = l_m \cdot A_c \left(\underbrace{\int_{-B_m}^{B_m} H \cdot dB}_{\text{Up-swing path}} - \underbrace{\int_{-B_m}^{B_m} H \cdot dB}_{\text{Down-swing path}} \right) f \tag{7.74}$$

It can be observed from Eq. (7.74) that the hysteresis power loss increases with the frequency of operation. The power loss also depends on the flux density and magnetization. However, it varies from core material to material as it is a function of the *B–H* loop path. It can be observed that if the up-swing path and the down-swing path coincide, then the hysteresis loss is zero as expected.

When the applied magnetic field *H* is removed (i.e., when the primary excitation is removed), the core flux is non-zero. The flux density can be anywhere either on OC part of axis or on OF part of axis, that is, positive or negative remnant flux density can exist. Such an uncertain *residual flux density* can create surge currents during switching ON of the transformer; therefore, one must accordingly account for this issue in the circuit design by incorporating soft-start circuits.

Eddy Current Loss

Whenever the flux in the core changes, that is $d\phi/dt$ is non-zero, a voltage is induced in the windings of the core according to Faraday's law. Observe that the core itself acts as a one turn winding (a closed turn) in addition to the other windings wound on the core.

Therefore a voltage is induced in the core which will cause a current to flow in the core material depending on the resistance offered by the core material. The eddy current i_e is a sheet of current that is distributed throughout the magnetic path length. The voltage induced for the single turn core material is

$$e_{eddy} = N \frac{d\phi}{dt} = \frac{d\phi}{dt} \quad (\text{as } N = 1)$$

The resistance offered by the core material for the path of the eddy current is

$$R_{eddy} = \frac{\rho l_e}{A_e} \tag{7.75}$$

where ρ is the resistivity of the core material, l_e is the mean length of the eddy current path and A_e is the mean cross-sectional area perpendicular to the flow of the eddy current. It should be noted that A_e is not the same as the core cross-sectional area A_c. The eddy current is given as

$$i_e = \frac{e_{eddy}}{R_{eddy}} \tag{7.76}$$

The power dissipated in the resistance of the core material is lost forever and is called the eddy current loss. This loss is given as

$$P_{eddy} = i_e^2 R_{eddy} = \frac{e_{eddy}^2}{R_{eddy}} = \frac{(d\phi/dt)^2 \cdot A_e}{\rho \cdot l_e} = \frac{4 \cdot f^2 \cdot A_c^2 \cdot \Delta B^2 \cdot A_e}{\rho \cdot l_e} \tag{7.77}$$

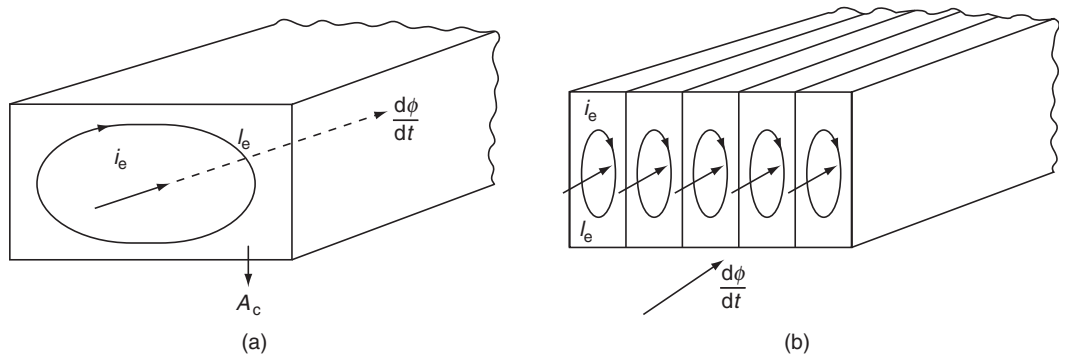


Figure 7.26 | Eddy current (a) in the core; (b) in a laminated core.

ΔB is the flux density swing in an interval of half a period. For a core where the flux swing is bi-directional like in power transformers, half- and full-bridge transformers, $\Delta B = 2B_m$. For cores where the flux swing is uni-directional like in inductors and forward converter or flyback topologies, $\Delta B = B_m$. Thus

$$P_{eddy} = \frac{8 \cdot f^2 \cdot A_c^2 \cdot B_m^2 \cdot A_e}{\rho \cdot l_e} \quad \text{(for bi-directional flux density swing)} \tag{7.78}$$

$$P_{eddy} = \frac{4 \cdot f^2 \cdot A_c^2 \cdot B_m^2 \cdot A_e}{\rho \cdot l_e} \quad \text{(for unidirectional flux density swing)} \tag{7.79}$$

It is difficult to estimate l_e and A_e for various core geometries. However, the core manufacturers provide nomo-graphs of the eddy current loss as a function of frequency for specific core geometries.

With some core materials like the silicon steel CRGO and CRNGO cores the material resistivity is very less. To reduce the resistance offered to the eddy current, the core cross-section is split into many small cross-sections called laminations as shown in Figure 7.26(b). This way $d\phi/dt$ also splits up by the number of laminations used. A_e is also reduced to increase R_{eddy}. If n laminations are used then

$$\frac{d\phi_l}{dt} = \frac{d\phi / dt}{m}$$

Thus,

$$P_{eddy} = m\left(\frac{4 \cdot f^2 \cdot A_c^2 \cdot \Delta B^2 \cdot A_{el}}{m^2 \cdot \rho \cdot l_{el}}\right) = \frac{4 \cdot f^2 \cdot A_c^2 \cdot \Delta B^2 \cdot A_{el}}{m \cdot \rho \cdot l_{el}} \tag{7.80}$$

where A_{el} is the mean cross-sectional area in a lamination perpendicular to the flow of the eddy current, l_{el} is the mean path of the eddy current in a lamination and m is the number of laminations.

As can be observed, the effective resistance of the core increases and the eddy currents reduce, thus, reducing the losses and temperature of the core. A second way of increasing the resistance of the solid core is to have material with large resistivity, such as ferrites and amorphous glass. They are a solid mass and laminations are not required due to their high inherent resistivity.

The hysteresis and eddy current losses together are called the *core losses*:

$$P_c = P_{hys} + P_{eddy} \tag{7.81}$$

Skin Effect

The current density across a conductor cross-section is not uniform. It varies from the conductor surface to the center of the conductor in an exponential manner. Figure 7.27 shows an enlarged view of a conductor

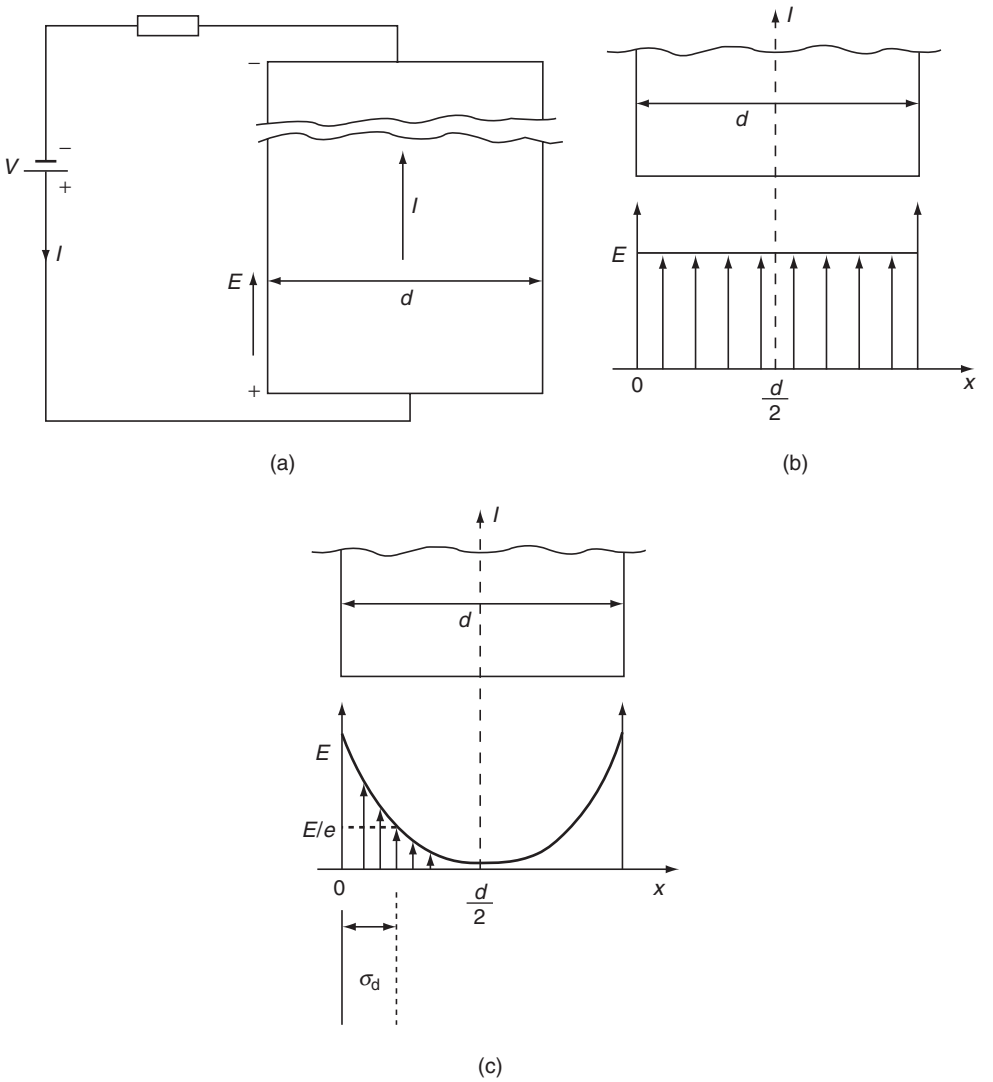


Figure 7.27 (a) *E*-field applied to a conductor; (b) DC *E*-field profile across the conductor width; (c) AC *E*-field profile across the conductor width.

carrying a current I. The conductor has a diameter d. If the current flowing through the conductor is DC, then the field E that acts as the motive force for the charges within the conductor is also DC. Figure 7.27(b) shows the distribution of the E-field across the width of the conductor. For a DC-field the E-field distribution across the width of the conductor is uniform as shown. This results in a uniform current distribution across the conductor cross-sectional area.

However, if the current is an AC, then the applied field in the conductor is also AC. The AC E-field has a maximum value at the conductor surface and decreases exponential towards the center of the conductor. The E-field distribution across the width of the conductor is shown in Figure 7.27(c) for an AC E-field. As the field strength near the surface of the conductor is high, the current density near the surface is higher than the central part of the conductor. This effect is called the *skin effect*.

The depth to which the E-field penetrates across the width of the conductor is measured by means of the parameter called the *skin depth*. *The skin depth is defined as the distance across the width of the conductor from the conductor surface to the point where the current density is 1/e times the surface current density*. It is denoted as σ_d. The skin depth is given as

$$\sigma_d = \sqrt{\frac{\rho}{\pi \cdot \mu_o \cdot \mu_r \cdot f}} \text{ m} \tag{7.82}$$

where ρ is the resistivity of the conductor material; μ_o the permeability of air; μ_r the relative permeability of the conductor; f the frequency of the current Hz.
For copper at 100°C, the resistivity is 2.3×10^{-8} Ωm and relative permeability is 1. Thus for copper

$$\sigma_d = \frac{7.6}{\sqrt{f}} \text{ cm} \tag{7.83}$$

It can be observed that the skin depth is dependent on the frequency. As the frequency increases, the skin depth reduces implying that at sufficiently high frequency the central portion of the conductor will not conduct any current and all the charges are concentrated near the surface of the conductor. This will decrease the effective cross-sectional area and increase the conductor resistance. As a consequence, the increase in the current density near the surface will increase the copper loss and heat up the conductor.

This problem can be reduced by splitting the conductor cross-section in multiple section conductors, that is, using multi-strand conductors wherein each strand is of a cross-section much small than the skin depth. In such a case the E-field will penetrate across the width of the conductor and the field distribution across the width will be more uniform. As a consequence, the conductor cross-section is better utilized resulting in lower copper losses. For very high frequency applications in excess of 200 kHz, the transformer windings are wound with multi-strand wires (like litz wires) to reduce the copper loss due to skin effect.

Proximity Effect

Consider two conductors wherein the currents are flowing in opposite directions. Figure 7.28(a) shows two conductor cross-sections wherein an "×" indicates current flowing into the page and "•" indicates current flowing out of the page. Due to the current flowing in the conductors the H-field pattern around each conductor is as indicated. The far field around both conductors has a tendency to cancel each other and the H-field between the conductors strengthens. If the two conductors have diameters that are greater than the skin depth at the frequency of operation, then it would be expected that the current density be higher near the surface as shown in Figure 7.28(b). However, the high-frequency current charges accumulate along one side of the surface to minimize the magnetic field energy (nature's minimum energy principle) and minimize

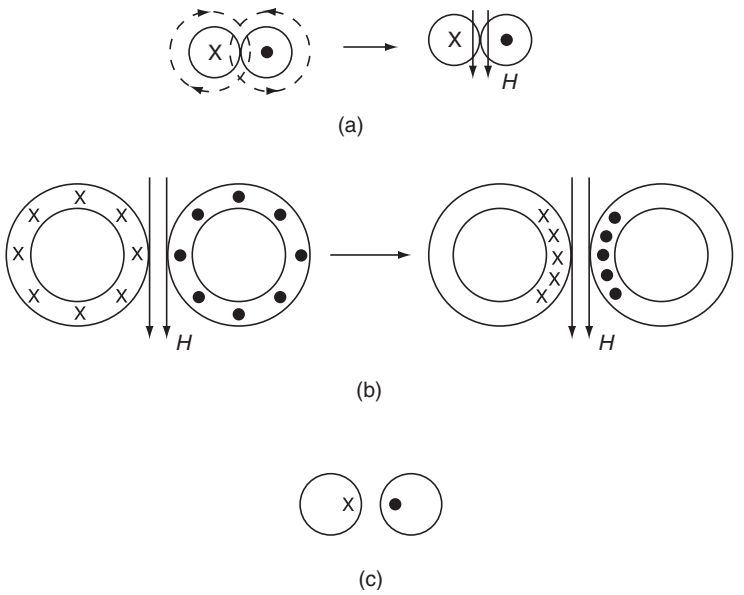


Figure 7.28 (a) *H*-field due to two conductors carrying equal and opposite currents; (b) charge accumulation due to skin and proximity effects; (c) proximity effect representation.

the inductive storage. Therefore, the current charges accumulate near the surface of the conductors opposite to each other as shown in Figure 7.28(b). This effect is called the proximity effect. Proximity effect and skin effect together makes the effective cross-sectional area of the conductor very small, leading to high copper losses. Figure 7.28(c) shows the symbolic representation of the proximity effect.

Effect Due to Unequal Currents

The proximity effect due to equal currents was discussed above. However, the proximity effect due to unequal currents has a much greater effect on the copper losses. The situation of unequal currents in the windings is more common than the equal current situation discussed above. Consider two conductors: one primary winding conductor and another secondary winding conductor that are in close proximity as shown in Figure 7.29. Let an "×" indicate a current of magnitude I flowing into the page and "•" indicate a current of magnitude I flowing out of the page. The transformer is designed such that the primary current carries a current of I and the secondary current carries a current of $2I$. If the frequency is high and the conductors are thicker than the skin depth at that frequency then the current charges in the primary and the secondary will accumulate at the surface of the conductors facing each other as shown in Figure 7.29. There are two dark dots shown in the secondary conductor representing a current of $2I$. There is one dark "×" mark shown in the primary conductor representing a current of I flowing in the conductor. Due to the proximity effect, the charges are accumulated close to each other near the surface of each conductor. The far field due to one dot of the secondary and one "×" of the primary will cancel; however, the other dot of the secondary will produce an H-field around it and induce a current of magnitude I to flow in the primary. This is marked as a light "×" mark. This induced current will circulate through the outer periphery as shown in the figure. This is shown as a light dot mark on the outer periphery. It should be observed that the total current through the

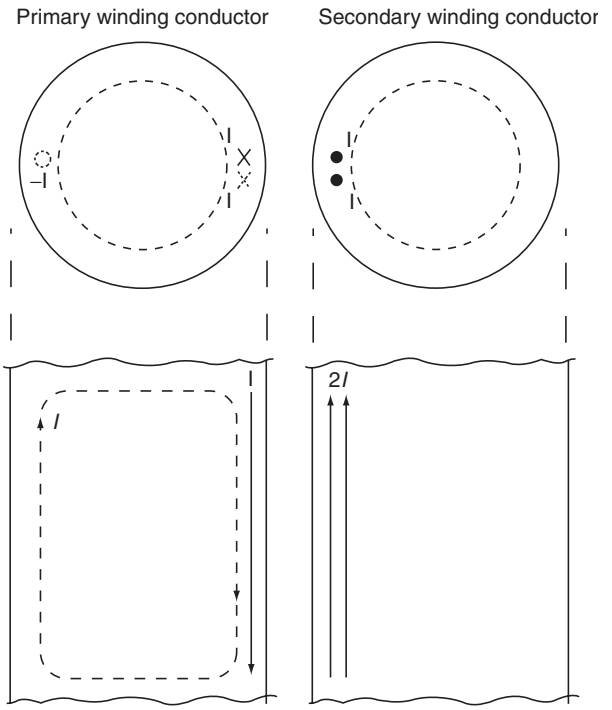


Figure 7.29 | Proximity effect due to unequal currents in conductors.

primary is still maintained as I. However, the I^2R loss is contributed due to all the current components flowing in the primary. Thus the loss is $(3I_{rms})^2R$ as against $I_{rms}{}^2R$ without proximity effect. Thus there is a nine times increase in the copper loss for twice the current in the secondary as that in the primary. With multiple layers, this has a further multiplying effect.

Thus proximity effect along with skin effect can be dangerously detrimental to the transformer especially at high-frequency operation. The solution is to use multi-strand wires and inter-layering the primary and secondary windings.

CONCLUDING REMARKS

Inductors and transformers are ubiquitious in any system, especially in power electronics systems. A thorough understanding of the magnetic domain principles is central to analysis and design of power electronics circuits. As discussed in this chapter, the Ampere's Law and the Faraday's Law are two fundamental laws that lay the foundation for the understanding of the magnetic domain aspects. These two laws form the framework for modeling of magnetic components.

The area product approach is widely followed in the design of the inductors and transformers. It is a very systematic and algorithmic approach that is easily amenable for computer-aided design and analysis. However, one should note that there are a couple of parameters that need to be decided by engineering judgment. One of the parameters is the current density J. Throughout the chapter, a default current density of 3 A/mm^2 has been used in the

examples. This seems to be a good compromise between the copper loss at high current densities and the large core size requirement at low current densities. However, the current densities can be selected in the neighborhood of the default value of 3 A/mm^2 depending on the application. In very low power applications, the current densities as high as 5–6 A/mm^2 have also been used with consequent increase in copper loss. In high-power applications, one can go as low as 2 A/mm^2 so that significant savings in the copper loss offsets the increase in the core size. The other variable that is left to engineering judgment is the window utilization factor. Though this is in the range of 0.3 to 0.6 based on the factors discussed in this chapter, it is also dependent on the skill with which the windings and the inter-winding insulation are incorporated.

The skin effect and the proximity effects are predominant in circuits where the switching frequency is in excess of 100 kHz. These effects will have a drastic effect of bringing down the efficiency of the system nullifying the primary objective of high-frequency switching which is to reduce the size of the magnetic components. Multi-strand wires are probably the best solution to reduce the effects of skin and proximity effects.

One more frequency-dependent effect that needs to be borne in mind is the core loss, that is, hysteresis and eddy current losses. Both increase with frequency. To reduce the hysteresis loss at high-frequency operation, first core materials with narrow hysteresis loop should be selected. Second, one may reduce the maximum operating flux density to a much lower value to reduce the *B–H* loop area in a cycle.

LABORATORY EXERCISES

1. Consider the inductor measurement circuit shown in Figure 7.30. The inductor *L* is the test inductor that is connected to the drain of a metal oxide semiconductor field effect transistor (MOSFET) as shown. A diode–resistor combination is used as the inductor current freewheeling circuit.

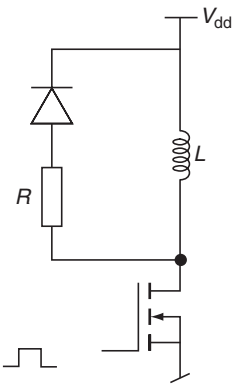


Figure 7.30 | Inductor measurement circuit.

Mode of implementation: The above circuit can be studied by

- **a.** Simulation in Spice
- **b.** Simulation in SciLAB
- **c.** Hardware bread-boarding

Tasks for study:

(a) Select a ferrite core from the stores or buy one (say E 42/21/15). From the datasheet of the core material find out the relative permeability, mean magnetic length and the core cross-sectional area. Calculate the permeance of the core. Wind 10 turns on the core bobbin and mount the bobbin onto the core. For the calculated permeance and the number of turns used for the coil, estimate the inductance.

(b) Rig up the circuit as shown in Figure 7.30 with appropriately selected component values.

(c) Connect a pulse generator and appropriate drive circuit to the gate of the MOSFET. Set the switching frequency to 20 kHz and keep the duty cycle close to 0. Increase the duty

cycle slightly and simultaneously observe the voltage across the MOSFET and the current through the inductor on an oscilloscope. When the switch is ON the current through the inductor will be a linearly rising waveform. Measure the slope of the inductor current waveform when the switch is ON. Also measure the voltage across the inductor. From Faraday's Law estimate the inductance value. Compare with the calculated inductance value of Step (a).

(d) Continue increasing the duty cycle till the current through the inductor starts to shoot up. Why does this happen? Observe the voltage across the inductor.

(e) Increase the number of turns in the coil to 20. What is the effect on the inductance?

(f) Introduce an air gap of 0.25 mm in the central limb of the E core. Calculate the permeance and the inductance. What is the effect of air gap on the inductance? Experimentally measure the inductance as in Step (c) and compare with the calculated value.

(g) For different values of the duty cycle, the peak inductor current values will be different. Tabulate the inductor current peak values at different duty cycles and estimate the magnetization field H.

2. Consider the transformer test circuit shown in Figure 7.31. The transformer with terminals A, B, C and D is the test transformer. Let AB be considered as the primary and CD as the secondary. V_i is an input source of an arbitrary waveshape and R_L is variable load resistor.

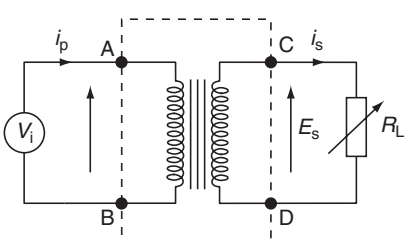


Figure 7.31 | Transformer test circuit.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Simulation in SciLAB
c. Hardware bread-boarding

Tasks for study:

(a) Take an unknown and unmarked transformer (low-frequency lamination type) and rig up the circuit as shown in Figure 7.31. Before rigging up the circuit, measure the physical dimensions and estimate the core cross-section area and the mean magnetic path length.

(b) The input voltage source is obtained from the 230 V grid. An auto-transformer is connected to the grid to obtain a variable sinusoidal voltage from the auto-transformer tap point to ground. Set the auto-transformer tap point to ground. While connecting the load resistor, change the auto-transformer tap point slowly to obtain 5 V. Measure the voltages V_{AB} and V_{CD} on the oscilloscope. Determine the dot polarities of the primary and the secondary windings. Determine also the turns ratio.

(c) Take a very low copper gauge wire, say SWG 45 and thread through the transformer about 10 turns. Apply a very low voltage to this 10-turn winding using the auto-transformer variable source. Measure the voltage across the 10-turn winding and the voltages across AB and CD. Estimate the number of turns in the AB winding and the number of turns in the CD winding. Remove the temporary 10-turn winding.

(d) From Eq. (7.50b) estimate the voltage that may be applied to AB. (As the laminations is of either CRGO or CRNGO material, assume B_m of 1 T for now.)

(e) Without connecting the load and by varying the auto-transformer setting, apply half the voltage to AB that is estimated in Step (d). Measure the primary voltage V_{AB} and the primary current through the winding AB. V_{AB}/N_{AB} gives the flux rate

within the core. Build an integrator circuit using op-amp, sense V_{AB} by using an isolation transformer or by resistive attenuation and integrate the ratio V_{AB}/N_{AB} to obtain the flux within the core. Divide this output by the core cross-sectional area to obtain the flux density within the core.

(f) From the measured primary current, estimate the *mmf* in the core by multiplying the number of turns in AB winding and the current through it. Sense the current through the winding by using a CT or by measuring the drop across a 0.1 Ω resistor connected in series with the winding. The voltage equivalent of the sensed current is multiplied by N_{AB}/l_m to obtain the magnetization field *H*.

(g) As V_{AB} is varying sinusoidally and the primary current is also varying sinusoidally, the flux density obtained in Step (e) and the magnetization field obtained in Step (f) are also AC waveforms. With these two waveforms, plot the *B–H* characteristics. Some oscilloscopes have "integrate" facility. In that case, the scope channel measuring V_{AB} can be integrated in the scope itself and the current can be given to the external trigger to see the scaled *B–H* characteristic shape on the oscilloscope.

(h) Increase the auto-transformer voltage slowly till the *B–H* curve just enters into saturation region. Tabulate the maximum B_m and the magnetization field at maximum flux density. From the *B–H* plot estimate the relative permeability of the core in the linear region. After estimating the relative permeability, calculate the permeance of the core. Calculate the mutual inductance of the core. Estimate the *B–H* loop area and calculate the core loss at the operating frequency. Tabulate the core loss at various input voltage (by varying the auto-transformer setting).

(i) Short circuit the secondary and measure the inductance as seen at AB. This will give the sum of the primary leakage and the reflected secondary leakage in parallel with the mutual inductance.

(j) Short circuit the primary and measure the inductance as seen at CD. This will give the sum of the secondary leakage and the primary leakage in parallel with the mutual inductance that is reflected to the secondary.

(k) Measure the winding diameters (using screw gauge) and estimate the winding cross-section area for the primary and secondary windings. Measure the resistances of the winding. From the resistivity of copper, determine the length of the copper wire used for the primary and the secondary windings.

(l) Based on the above measurements, construct the equivalent circuit for the test transformer.

(m) What is the effect of variation in load resistor R_L on the *B–H* characteristic and the equivalent circuit parameters?

FILL IN THE BLANKS

1. The magnetics components are categorized as energy-_____ devices and energy-_____ devices.

2. The energy-_____ devices are called the transformers.

3. The energy-_____ devices are called inductors.

4. The flux density is related to the flux as _____.

5. The power associated with the saturated core in the magnetic domain is _____.

6. The volt-second product for a coil in terms of the physical parameters of the core is given as _____.

7. For a transformer of a given rated voltage, if a lower frequency is applied, the _____ will increase and may saturate the core.
8. The voltage across a coil should have a _____ average over a full cycle.
9. In an inductor, the energy is stored by virtue of the _____ flowing through the inductor in the electrical domain.
10. The energy-storage mechanism within the magnetic domain is a _____ storage similar to that of _____ storage in electric domain.
11. The permeance is equivalent to a _____ within the magnetic domain.
12. The core cross sectional area is related to the _____ across the coil.
13. The window area of the core is related to the _____ capacity.
14. The area product relates to the _____ that is stored in the core or the _____ that is transferred through the core.
15. The flyback transformer is an energy-_____ device. It is nothing but a multiple-winding _____.
16. If the power source in the primary of a transformer is a voltage source, then the transformer is called a _____.
17. If the power source in the primary of a transformer is a current source, then the transformer is called a _____.
18. The primary current needed to supply is core losses and to set up the magnetic flux is given as _____.
19. In the equation, $E_{\text{p-av}} = 2 \cdot N_p \cdot A_c \cdot \Delta B \cdot f$, the variable $E_{\text{p-av}}$ is the _____ average of the induced voltage across the primary winding.
20. Current transformers are mainly used to _____ the currents in the power circuits.
21. Short circuiting the secondary winding in the case of the current transformer implies _____ operation.
22. As CTs are used as current sensing devices, the main focus is on maintaining the current waveshape _____.
23. Rod cores are mostly used in radios for _____ an inductor.
24. In rod core type of inductor, radiation occurs at the end of the rod and will lead to generation of _____ radiation.
25. The pot core has a shielding effect, preventing radiation and _____ electromagnetic interference.
26. Toroidal cores are used in _____ transformers that are popular in amplifiers, but it is not easy to _____ a coil on toroidal cores.
27. The hysteresis power loss _____ with the frequency of operation.
28. The current density near the centre of the conductor is _____ than that near the surface of the conductor due to skin effect.

DESCRIPTIVE QUESTIONS

1. What are the two fundamental laws of electromagnetism that link voltage and current in the electrical domain to equivalent quantities in the magnetic domain?
2. What is Ampere's Law?
3. What is Faraday's Law?
4. To what equivalent variables do voltage, current and charge of the electric domain map to in the magnetic domain?

5. What is the equivalent of resistance in the magnetic domain?
6. What is the equivalent of capacitance in the magnetic domain?
7. Distinguish between resistance, capacitance, permeance and reluctance.
8. How are magnetic field intensity and flux density related?
9. Explain the *B*–*H* characteristics and the hysteresis concept.
10. Explain the volt-second balance principle.
11. What is the effect on the *B*–*H* characteristic of a core when an air gap is introduced?
12. Discuss the window utilization factor.
13. Explain how the flux rate within the core is maintained constant.
14. Explain how the change of load in the secondary winding gets reflected on the primary side.
15. Under no-load operation, is there any current drawn from the input source connected to the primary? Why is this current needed?
16. What is turns ratio? Discuss.
17. What is leakage flux? Explain.
18. Explain the equivalent circuit of the transformer.
19. What are the features of an ideal transformer and how does a practical transformer deviate from ideality?
20. Explain the differences between the current transformer and the potential transformer.
21. In what ways can the magnetizing current be reduced in a CT?
22. What are the physical parameter requirements for a good quality CT?
23. What are the factors that affect the utilization of the window area of a core? Discuss.
24. Discuss the various shapes of the commercial cores illustrating the core cross-sectional area and the window area in the front and profile views of a core section?
25. Explain how one would test the dot polarity of an unknown and unmarked transformer.
26. Discuss the hysteresis loss in the cores. What is the effect of frequency on the loss?
27. What are the factors affecting the hysteresis loss in a core?
28. What is eddy current loss and how can this be minimized?
29. What are the factors affecting the eddy current loss?
30. Explain the skin-effect phenomenon. How is the effect of skin effect reduced in the windings of high-frequency transformers?
31. Explain the proximity effect. In what ways can the proximity effect be reduced?

PROBLEMS

1. A ferrite EE core having a cross-section area of 182 mm^2, has a mean magnetic path length of 97.2 mm. If the permeance is 4.7 $\mu H/turns^2$, find the reluctance and the relative permeability of the core.
2. An inductor core has a permeance of 4.7 $\mu H/turns^2$. The magneto-motive force within the core is 20 ampere-turns. What is the flux in the core?
3. For Problem 2, if the number of turns in the coil winding is 10, what is the inductance of the coil?
4. 100 turns of copper wire are wound on a cylindrical plastic bobbin in a single layer. The outer

diameter of the bobbin is 16 mm and the length of the coil is 50 mm. If no core is inserted into the bobbin, what is the inductance of the coil?

5. For Problem 4, if the bobbin is inserted in a tight fitting CC ferrite core of relative permeability 2000 having a mean magnetic path length of 150 mm, then what is the inductance of the coil?

6. An inductor core has a permeance of 10 μH/turns^2. The magneto-motive force within the core is 25 ampere-turns. What is the energy stored in the inductor?

7. For Problem 4, if a ferrite rod of length 75 mm is tightly fitted into the bobbin, then what is the inductance of the coil if the ferrite material has a relative permeability of 2000? Also calculate the reluctance of the magnetic path. (Hint: The flux path is through the ferrite rod and returns back through approximately an equal distance in air.)

8. A particular DC–DC converter that is switched at 50 kHz, uses a 100 μH inductor wherein 10 A of DC current flows through it. The ripple of 1 A peak to peak is superimposed on the 10 A DC current. Calculate the area product requirement for a ferrite cored inductor. Also select an appropriate core from Appendix II.

9. For a ferrite core E 42/21/9 whose physical parameters are given in Appendix II, calculate the permeance if the relative permeability of the core is 2000 and the introduced air-gap length is 0.5 mm.

10. If the required inductance is 100 mH and the permeance is as calculated in Problem 9 for the ferrite core E 42/21/9, estimate the number of turns required to be wound on the core.

11. For a 10 A DC current flowing through the inductor coil, calculate the wire cross-section that is required for the coil wire. Also select an appropriate gauge of the wire from Appendix III.

12. A transformer having 1000 turns in the primary is connected to voltage source whose waveshape is given as 100 sinωt. What is the waveform of the flux rate?

13. A transformer is designed to take power from an input source rated at 230 V rms, 50 Hz sinusoidal source. The transformer comprises two secondary windings that supply power to a 100 V, 2 A and 50 V, 5 A loads. Assuming CRGO silicon steel laminations for the transformer and an efficiency of 80%, estimate the required area product such that the maximum operating flux density in the core does not exceed 1 T? Select an appropriate lamination type from Appendix II.

14. For Problem 13, find the turns ratios and calculate the number of turns for the primary and the secondary windings.

ANSWERS

Fill in the Blanks

1. transfer; storage
2. transfer
3. storage
4. $B = \phi/A_c$
5. zero
6. $E_{avg}T = 2NA_cB$
7. $E_{avg}T$
8. zero
9. current
10. potential; capacitance
11. capacitance
12. voltage
13. current
14. energy; power
15. storing; inductor
16. potential transformer
17. current transformer
18. $i_{po} = i_p - n \cdot i_s$
19. half-period
20. sense
21. no-load
22. fidelity
23. tuning
24. electromagnetic
25. reducing
26. low radiation; wind
27. increases
28. lesser

CHAPTER 8

Modeling of Systems

Learning Objectives

After reading this chapter, you will be able to:

- make a mathematical representation of a physical system.
- obtain the state space model of a physical system.
- obtain the transfer function model of a physical system.
- compare and use various modeling approaches.
- apply circuit averaging, bond graph and space-vector modeling methods for power electronics systems.

Till now, the models used to describe the systems in terms of the input–output relationships of the converters were steady-state models. These models reflect the state of system under equilibrium conditions. However, the dynamics of the systems are important from the point of view of controller design for regulation and reference tracking. The most important task confronting the system analyst is developing the dynamic model of the process to be controlled. In most cases, the central problem in any design situation is modeling. Once the modeling is performed with a reasonable degree of precision, the rest of the analysis becomes more or less a trivial issue. Determination of a mathematical model of a given physical system is an important problem in engineering design and analysis. The model must relate the various variables in the system in a quantitative manner. A model may be defined as "a representation of the essential aspects of a system which presents knowledge of that system in a usable manner". To be useful, a model must not be so complicated that it cannot be understood and thereby be unsuitable for analysis; at the same time, it must not be oversimplified and trivial to the extent that predictions of the behavior of the system based on this model are grossly inadequate.

The control system engineer is most often required to deal with systems that have many sub-systems operating in different physical domains. The electromechanical process, for example, may comprise the DC motor coupled to inertia load on one side and electrical source on the other. The electrical source itself will have its associated dynamics and the mechanical inertia load would have its own dynamics that needs to be controlled. Further, the DC motor itself includes the magnetic domain with associated dynamics. Therefore, for the control system engineer to design a practical meaningful controller it is required to first model the entire system with all its different domains and physical laws such that the system is represented in a mathematical form that is amenable for analysis, computation, simulation and design.

As the control system engineer has to deal with systems having multi- and interdisciplinary domains, it implies that one has to have a deeper understanding of the various sub-systems that go to make up the whole system. Therefore, for a control systems engineer, knowledge of various domains is very important. The information about the various sub-systems of the different domains should be translated into a form that is compatible for controller synthesis. The equations and formulae that predict the dynamic behavior of the system as a whole should be established. The behavior of the overall system can then be analyzed and characterized so that a controller may be designed to meet the desired performance specifications.

There are basically two modeling approaches: (a) system identification and (b) physical modeling. System identification is generally associated with automated computer-aided modeling. The system is considered as a black box, the output of which is characterized for various inputs. Various inputs like sinusoidal, step, etc. are applied to the actual system. The response to these various types of inputs are characterized as transfer curves to obtain the input–output relationships. Based on the transfer curves, a mathematical curve fit algorithm is proposed such that it gives the best description of the system. System identification technique is generally used in on-line estimation of system parameters like time constants, resistances, etc. for self-tuning and adaptive controllers.

The physical modeling techniques, on the other hand, obtain the mathematical representation of the physical system from first principles. This provides a much deeper insight into the behavior of the various dynamics of the system as compared to the system identification methods. The system is represented as a set of differential equations that describe the dynamic behavior of the system. For designing controllers it is essential to obtain the mathematical representation from first principles so that it will aid in the analysis and synthesis of the system to be controlled together with the controller. This chapter focuses on obtaining the physical model of systems from first principles in the differential equation form.

The physical systems that are considered for analysis are dynamic in nature, and their behavior will be described in the form of differential equations. Although these will normally be non-linear, it is customary to linearize them about an operating point to obtain linear differential equations. This is done for the analysis to be carried out conveniently. It should, however, be noted that such linear models, though useful for analysis and design, are valid only over a limited operating range. Nevertheless, they are employed extensively in engineering.

The central issue in any control system analysis or design study is to first obtain a reasonable mathematical representation of the physical system. In obtaining a mathematical model of a given physical system, one must make a compromise between simplicity of representation and accuracy of the prediction of the system behavior. In deriving a model, it may be necessary to ignore certain physical properties of the system such as distributed capacitances and inductance, non-linearities such as saturation effects that may be present in the system. Only those properties that do not have a significant effect on the response should be removed. This will lead to simpler mathematical models having reasonable agreement with the experimental study.

In general, it is important to note that one should build a simplified model of the system so that an intuitive understanding of the behavior of the system is possible. For more complete and precise analysis, a more detailed model can later be evolved from the simplified model. It should also be borne in mind that while one can make reasonable approximations to obtain simplified models of the system, the simplified models may not be valid at other operating points. For example, if the distributed capacitances and inductances are replaced by lumped parameter approximations, the model though valid at low frequencies may not be valid at high frequencies. Therefore for every model, it is important to understand its limitations with respect to the range of operation.

8.1 Input–Output Relations

Consider a system which is shown as a black box in Figure 8.1. It has an input u and an output y as indicated. The aim of modeling is to predict the output y for any given input u. The output y is called the response of the system and u is called the stimulus or simply the input to the system.

The input–output relations for the system may be given in a graphical form called input–output transfer curves or input–output characteristic curves as shown in Figure 8.2. These curves could be obtained by plotting the experimental response obtained for various input signal values. The curves are plotted with the input u as the independent variable (x-axis) and the response y as the dependent variable (y-axis).

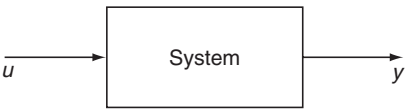


Figure 8.1 | System as a black box.

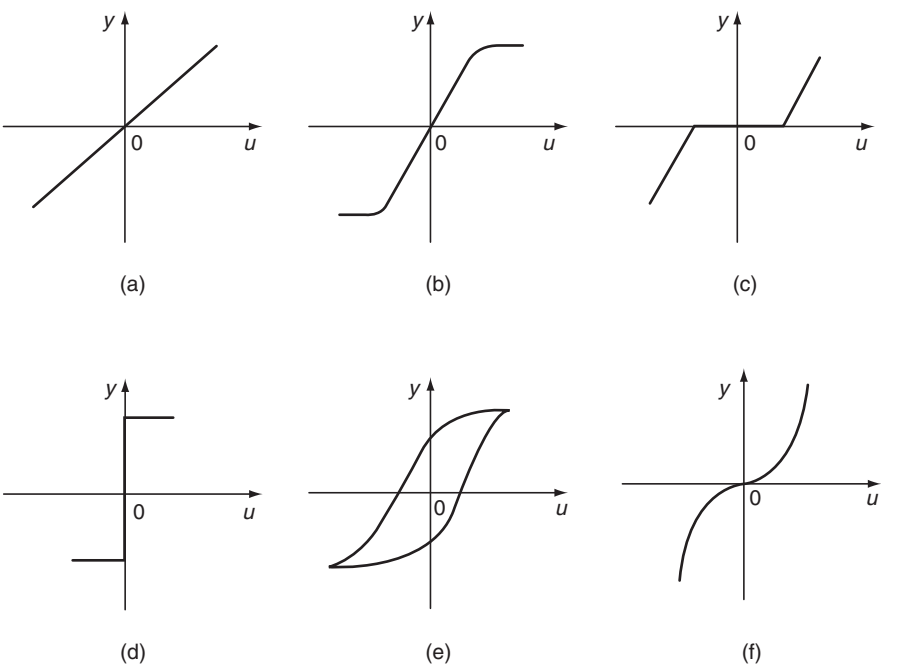


Figure 8.2 | Different types of input–output transfer curves: (a) Linear; (b) non-linear saturation effect; (c) non-linear dead zone effect; (d) non-linear switch; (e) non-linear hysteresis; (f) non-linear exponential law.

Figure 8.2(a) shows the input–output transfer curve of a linear system wherein the response is proportional to the input. It should be noted here that in a linear system the most important property is that the principle of superposition is applicable. This means that the response produced by the simultaneous application of two different stimuli is the sum of the two individual responses. A system wherein the principle of superposition is not applicable is called a non-linear system. Figures 8.2(b)–(f) give some typical input–output transfer curves for non-linear systems. In Figure 8.2(b), one can see the saturation effect wherein the response y varies linearly with u upto a point. Beyond some value of u, the system output clamps to a specific value. Saturation is very typical of most physical systems. Figure 8.2(c) is an example of dead zone non-linearity wherein at very low values of the input signal u, there is no response from the system. The system will respond only if an input above a threshold value is applied. Figure 8.2(d) is an example of a switch. Here the output is a two-state system wherein it will be high for inputs that are greater than zero or a specified threshold value and the output will be low for inputs that are lower than zero or a specified threshold value. Figure 8.2(e) is an example of hysteresis non-linearity. Here the output follows one path for increasing input u and follows another path for decreasing input u. This is typical of many dynamical systems. The B–H curves of magnetic materials are a typical example of hysteresis non-linearity. Figure 8.2(f) shows an example of exponential non-linearity wherein $y = e^u$. Similarly, the output can also follow square law, hyperbolic law, etc.

The input–output relationships can also be represented in the form of equations. Some examples are as follows:

1. **Linear System:** $y = ku$ where k is a constant.
2. **Non-Linear System:**
   - $y = u^2$. This input–output relation shows square law dependency.
   - $y = \cos u$. This input–output relation shows cosine dependency.

8.2 Differential Equations and Linearization

The input–output transfer curves do not give any information on the time evolution of the output to a given system input. When an input u is applied to the system, the output will finally settle down to a value as given by the input–output transfer curve. However, one is also interested to trace the time evolution of the output to its final settling value on the application of an input. This is called the dynamic behavior of the system and is represented by a set of differential equations. In practice, some simplifying engineering assumptions are often made to obtain linear differential equations with constant coefficients, although in most cases exact analysis would require the use of non-linear partial differential equations.

Consider a system with input $u(t)$ and output $y(t)$. Let it be a non-linear system with the input–output transfer curve as given in Figure 8.3. The input–output relationship is given by

$$y = f(u) \tag{8.1}$$

Referring to Figure 8.3, the nominal input to the system is $\bar{u}$ and the resulting output is $\bar{y}$. Therefore, $(\bar{u}, \bar{y})$ is the nominal operating point of the system. In a very small neighborhood about the operating point given by $(\Delta u, \Delta y)$, the system operation can be considered to be linear. For non-linear systems, the linear differential equations will be obtained for a very small neighborhood about the operating point such that the system can be considered to be linear in that small neighborhood about the operating point.

Equation (8.1) can be expanded into a Taylor series about the nominal operating point $(\bar{u}, \bar{y})$ as

$$y = f(u) = f(\bar{u}) + \frac{\mathrm{d}f}{\mathrm{d}u}(u - \bar{u}) + \frac{1}{2!}\frac{\mathrm{d}^2 f}{\mathrm{d}u^2}(u - \bar{u})^2 + \cdots \tag{8.2}$$

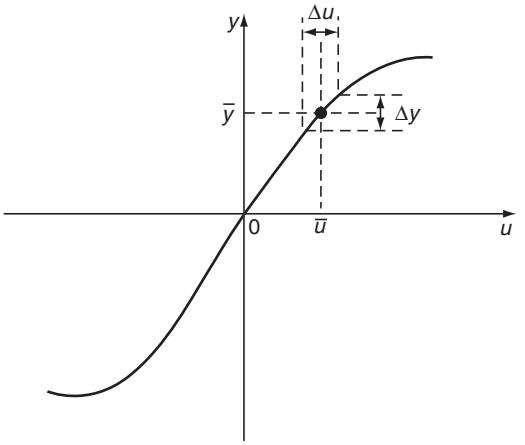


Figure 8.3 | Operating point and neighborhood of operating point.

In Eq. (8.2), the derivatives df/du, d^2f/du^2, etc. are evaluated at $\bar{u}$. If the deviation $(u-\bar{u})$ is small, then the second- and higher order terms may be neglected. Equation (8.2) may be simplified as

$$y = \bar{y} + k(u-\bar{u}) \tag{8.3a}$$

where

$$\bar{y} = f(\bar{u}) \tag{8.3b}$$

and $k = df/du\big|_{u=\bar{u}}$ which may be considered to be a constant in the neighborhood about the operating point. Equation (8.3a) may be re-written as

$$\Delta y = k \cdot \Delta u \tag{8.4}$$

where

$$\Delta y = y - \bar{y}$$
$$\Delta u = u - \bar{u}$$

Equation (8.4) shows that the deviation in the output is proportional to the deviation in the input and is the linearized mathematical model of the non-linear system given by Eq. (8.1). Equation (8.1) is called the *large-signal model* of the system. Equation (8.3b) is called the *steady-state or equilibrium model* of the system. Equation (8.4) is called the *small-signal model* of the system which is obtained by subtracting the steady-state model from the large-signal model.

Control system deals with components which are diverse in nature and may include electrical, mechanical, thermal and hydraulic systems. The differential equations for these devices are obtained using the basic laws of physics. For most physical systems one may classify the variables as "through" and "across" variable. The "through" variable can also be described as "flow" variable and the "across" variable can also be described as the "effort" variable. A list of analogous variables for different systems is given in Table 8.1.

Table 8.1 | Analogous variables for physical systems

System	Through or flow variable	Across or effort variable
Electrical	Current, I	Voltage, e
Mechanical (translation)	Velocity, v	Force, f
Mechanical (rotational)	Angular velocity, ω	Torque, T
Thermal	Rate of flow of heat energy, q	Difference in temperature, ΔT
Hydraulic	Volumetric rate of flow of fluid flow, dQ/dt	Difference in pressure, ΔP

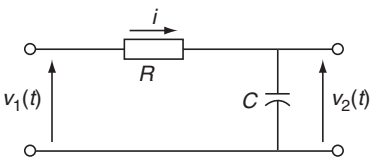


Figure 8.4 | A simple low-pass filter.

It may be noted that the equations of equilibrium in different systems are based on energy conservation principles. Kirchhoff's voltage law (KVL) for an electrical circuit, equating the algebraic sum of voltages in a loop to zero, is analogous to D'Alembert's principle in mechanics, which equates the algebraic sum of forces at a point to zero. These principles are used to obtain the differential equations characterizing the dynamics of the system.

For example, consider the simple electrical circuit shown in Figure 8.4, wherein an input voltage $v_1(t)$ is applied to an *RC* network. By the application of KVL to the circuit of Figure 8.4, the output and input are related as

$$\begin{aligned} v_1(t) &= R \cdot i + v_2(t) \\ v_1(t) &= R \cdot C \frac{dv_2}{dt} + v_2(t) \end{aligned} \tag{8.5}$$

Equation (8.5) represents the mathematical model of the *RC* low-pass filter of Figure 8.4. The solution to the differential equation given in Eq. (8.5) gives the prediction of the dynamic behavior of the *RC* low-pass filter.

8.3 State Space Representation

The notion of the state of a dynamical system is a fundamental concept which is central to the analysis of physical systems. The future evolution of a dynamical system is entirely determined by its present state. The state of a dynamic system may be defined as

> the smallest set of physical quantities called state variables that completely determines the evolution of the system in the absence of external excitation.

For any given physical system, there exists a set of physical quantities called state variables. For a system where the input is removed, the system behavior is entirely decided by the values of the state variables. The values of the state variables at any given point of time are called the state of the system. From the definition, it also implies that the values of the state variables at a given point of time along with the values of the inputs entirely dictate the future evolution of the state of the system.

The specific physical quantities (state variables) that define the system state are not unique. However, the number of state variables called the order of the system is unique for a given system. In many cases the choice of state variables is not very obvious. For a variable to qualify for consideration as a state variable, it should have the important property of memory. This means that it is capable of storing immediate past history. The following can qualify as state variables:

1. A memory unit of a digital system.

2. A variable that is representable in integral form. This is due to the fact that an integrator has memory effect. Inductor currents are representable in integral form as

$$i = \frac{1}{L}\int v \cdot \mathrm{d}t$$

Therefore an inductor current can qualify for consideration as a state variable. Capacitor voltages are representable in integral form as

$$v = \frac{1}{C}\int i \cdot \mathrm{d}t$$

Therefore, capacitor voltages can also qualify for consideration as a state variable. Based on similar argument, one can state that all the "flow" or "through" variables of inductor type or inertial type elements in any domain can qualify to be state variables. This means that velocity of a mass, angular velocity of inertia, etc. can qualify as state variables. Likewise, all the "effort" or "across" variables of capacitor type elements in any domain can qualify to be state variables. This means that force of a spring, temperature of a thermal capacitance, mmf of a magnetic permeance, etc. can qualify as state variables.

Another interesting point to be noted is that as the state variables should have memory property; they should be associated with energy-storing elements like inductors and capacitors only. Voltages across and current through resistors cannot qualify for being considered as state variables. In inductors the energy storage is due to current which is given by $(1/2)Li^2$ and in capacitors the energy is stored due to the voltage across the capacitors which is given by $(1/2)Cv^2$. Therefore, as a rule the variable responsible for energy storage in the devices can qualify for being considered as state variable.

For the *RC* low-pass filter circuit shown in Figure 8.4, only the voltage across the capacitor $v_2(t)$ can qualify for being a state variable. The state variable $v_2(t)$ along with the input $v_1(t)$ completely describes the dynamic behavior of the circuit of Figure 8.4. The differential equation [Eq. (8.5)] can be re-written as

$$\frac{\mathrm{d}v_2}{\mathrm{d}t} = \frac{-1}{RC}v_2 + \frac{1}{RC}v_1 \tag{8.6}$$

This is the standard form in which the system is represented and is called the state equation. This differential equation is entirely a function of state variables and the input variable. If there are n state variables then the system is said to be of order n. Such an order n system would be represented by n linear differential equations which are all functions of only the state variables and the input. Equation (8.6) can be further standardized as

$$\dot{x} = a \cdot x + b \cdot u \tag{8.7}$$

Equation (8.7) is called the state equation. With respect to the *RC* network circuit that is represented by Eq. (8.6), the parameters of the standard form of the state equation [Eq. (8.7)] can be related as

$$\dot{x} = \frac{\mathrm{d}v_2}{\mathrm{d}t}$$
$$a = \frac{-1}{RC}$$
$$b = \frac{1}{RC}$$

One can now explicitly also give the output as a function of only the state variables and the input variables. This can be given in a standard form as

$$y = c \cdot x + d \cdot u \tag{8.8}$$

Equation (8.8) is called the output equation. With respect to the *RC* network of Figure 8.4, the output equation is as given in Eq. (8.8) with the following parameters:

$$y = v_2$$
$$c = 1$$
$$d = 0$$

Equations (8.7) and (8.8) together represent the system in state space form. For more than one state, x becomes a vector and a, b, c and d become the matrices. Therefore, for an nth order system, there will be n state variables that define the state space.

Consider an *RLC* circuit as shown in Figure 8.5. The circuit is excited by an input v_i to obtain a response across the output v_o. Here the current (i) through the inductor and the voltage across the capacitor (v_o) qualify to be considered as state variables. Therefore, the entire *RLC* system will be completely described by the two state variables (i, v_o) and the input variable v_i. As there are two state variables, the order of the system is 2. One can expect two linear first-order differential equations as a function of the two state variables and the input variable. The two linear first-order differential equations that represent the *RLC* circuit are obtained using KVL. They are

$$\begin{aligned} v_\mathrm{i} &= iR + L\frac{\mathrm{d}i}{\mathrm{d}t} + v_\mathrm{o} \\ i &= C\frac{\mathrm{d}v_\mathrm{o}}{\mathrm{d}t} \end{aligned} \tag{8.9}$$

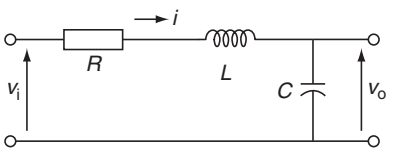


Figure 8.5 | *RLC* circuit.

Equation (8.9) can be re-written in the standard form as

$$\begin{aligned}\frac{\mathrm{d}i}{\mathrm{d}t} &= \frac{-R}{L}\cdot i + \frac{-1}{L}\cdot v_{\mathrm{o}} + \frac{1}{L}\cdot v_{\mathrm{i}} \\ \frac{\mathrm{d}v_{\mathrm{o}}}{\mathrm{d}t} &= \frac{1}{C}\cdot i + 0\cdot v_{\mathrm{o}} + 0\cdot v_{\mathrm{i}}\end{aligned} \tag{8.10}$$

Equation (8.10) can be re-written in the matrix form as

$$\begin{bmatrix} \mathrm{d}i/\mathrm{d}t \\ \mathrm{d}v_{\mathrm{o}}/\mathrm{d}t \end{bmatrix} = \begin{bmatrix} -R/L & -1/L \\ 1/C & 0 \end{bmatrix} \cdot \begin{bmatrix} i \\ v_{\mathrm{o}} \end{bmatrix} + \begin{bmatrix} 1/L \\ 0 \end{bmatrix} \cdot v_{\mathrm{i}} \tag{8.11}$$

Equation (8.11) is the state equation which represents the dynamic behavior of the *RLC* circuit of Figure 8.5. The output equation is given as

$$y = [0\ 1] \cdot \begin{bmatrix} i \\ v_{\mathrm{o}} \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix} \cdot v_{\mathrm{i}} \tag{8.12}$$

Equations (8.11) and (8.12) completely represent the *RLC* circuit in the state space form. Equations (8.11) and (8.12) are in the general form

$$\begin{aligned}\dot{x} &= A\cdot x + B\cdot u \\ y &= C\cdot x + D\cdot u\end{aligned} \tag{8.13}$$

For any system the mathematical model can be represented in the standard form of Eq. (8.13). Here x is the $n \times 1$ state vector which comprises the state variables for a system of order n, A is the $n \times n$ system parameter matrix which consists of the system constants. The matrix A contains the system-specific information about the dynamic behavior like the various time constants of the system. If the elements of A are constants of the system, then the system is called a linear time-invariant (LTI) system. If the elements of A are functions of time, then the system is called a linear time-varying (LTV) system. The matrix B is the $n \times m$ input matrix which weights the direct input excitation for state variable. U is the $m \times 1$ input vector for m input excitations, y is the $p \times 1$ output vector for p outputs from the system. C is a $p \times n$ output matrix and D is a $p \times m$ matrix that gives the direct feed through component of the input excitation in the output response.

8.4 Transfer Function Representation

Laplace transformation of the differential equations yields an algebraic equation, in terms of the complex frequency variable s, relating the various "through" and "across" variables. These algebraic equations can then easily be manipulated to obtain the transfer function of the system, defined as the ratio of the Laplace transforms of the output to the input under zero initial conditions. This transfer function represents a linear model of the system and is usually shown in the form of a block diagram as indicated in Figure 8.6. Note that the block is "unidirectional".

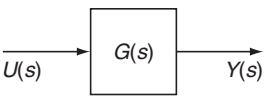


Figure 8.6 | Block diagram of a linear system.

The input may be regarded as the "cause" and the output as the "effect". The block diagram is unidirectional since the "effect" cannot produce the "cause". The transfer function $G(s)$ relates the Laplace transform $Y(s)$ of the output $y(t)$ to the Laplace transform $U(s)$ of the input $u(t)$ through the relationship

$$Y(s) = G(s)U(s) \tag{8.14}$$

For example, consider the *RC* electrical circuit shown in Figure 8.4. The output voltage $v_2(t)$ is related to the input through the differential equation as given in Eq. (8.5) which is repeated here for clarity:

$$v_1(t) = R \cdot C \frac{\mathrm{d}v_2}{\mathrm{d}t} + v_2(t)$$

Taking the Laplace transform on both sides of the above equation and assuming zero initial conditions, one obtains

$$V_1(s) = RCsV_2(s) + V_2(s) \tag{8.15}$$

From Eq. (8.15) and the definition of the transfer function, one obtains the transfer function as

$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{1}{1 + sRC} \tag{8.16}$$

The transfer functions are also basically derived from the differential equations. Therefore, the transfer function representation is derivable from the state space representation.

Consider the state space representation given in Eq. (8.13) which is repeated here for clarity:

$$\dot{x} = A \cdot x + B \cdot u$$
$$y = C \cdot x + D \cdot u$$

Taking the Laplace transform of the state equation yields

$$sX(s) - x(0) = AX(s) + BU(s)$$

or

$$(sI - A)X(s) = x(0) + BU(s)$$

Pre-multiplying both sides of the above equation by $(sI - A)^{-1}$, one obtains

$$X(s) = (sI - A)^{-1}x(0) + (sI - A)^{-1}BU(s) \tag{8.17}$$

From the definition of the transfer function, the initial conditions are zero. Equation (8.17) under the condition of zero initial condition is given as

$$X(s) = (sI - A)^{-1}BU(s) \tag{8.18}$$

From the output equation of the state space representation of Eq. (8.13), one obtains

$$Y(s) = CX(s) + DU(s) \tag{8.19}$$

From Eqs. (8.18) and (8.19), the transfer function from the state space representation is given as

$$G(s) = \frac{Y(s)}{U(s)} = C(sI - A)^{-1}B + D \tag{8.20}$$

For complex systems, it is easier to obtain the state space representation in the form of a set of linear first-order differential equations. This can later be used to obtain specific transfer functions for specific outputs with

respect to specific inputs. It should be noted that transfer functions are defined between a single input and a single output, that is, only for a single-input single-output (SISO) system. Therefore, for multiple-input and multiple-output (MIMO) systems one can represent the input–output relationship through a transfer function matrix. Each element of this matrix is a transfer function relating a particular output to a specific input, assuming that all the other inputs are zero. This is justified through the principle of superposition for linear systems.

For engineers with electrical background, the analysis of mechanical systems is often easier when an electrical equivalent circuit analogous to a mechanical system is obtained. It has the advantage that one can apply Kirchhoff's laws to write the circuit equations and hence obtain the transfer function. It is also possible to write these equations directly in terms of the Laplace transforms of the currents and voltages. Furthermore, network theorems can often be applied to simplify the circuit. Electrical analogs for mechanical systems have also been used for simulation and analysis. The rule for drawing the equivalent electrical circuit may be stated as follows:

Each junction in the mechanical system corresponds to a node in the equivalent electrical circuit, joining excitation sources and passive elements. All points on a rigid mass are considered as the same junction and one terminal of the capacitor analogous to the mass is always connected to the ground in the electrical circuit. The reason for connecting one terminal of the capacitor to the ground is that the velocity (or displacement) of a mass is considered with respect to a reference.
The following examples will illustrate the procedure.

EXAMPLE 8.1

The electrical analog of a carriage on wheels, coupled to the wall through a spring, is shown in Figure 8.7. The differential equation for both systems is as given in Eq. (8.21). In the case of the electrical network, the equation is obtained by applying Kirchhoff's current law (KCL) at the node and is seen to be identical to the equation that would have been obtained by equating the forces on all the components to the applied external force in the mechanical system.

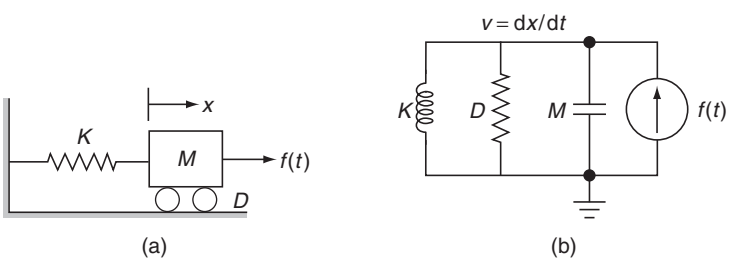


Figure 8.7 A mechanical system with movement in one co-ordinate: (a) Mechanical system; (b) electrical analog.

$$M\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + D\frac{\mathrm{d}x}{\mathrm{d}t} + Kx = f(t) \tag{8.21}$$

where M is the mass, K the spring stiffness, D the damper, x the displacement and f the applied force. Taking Laplace transforms of both sides of Eq. (8.21) and assuming zero initial conditions, one obtains the transfer function as

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{Ms^2 + Ds + K} \tag{8.22}$$

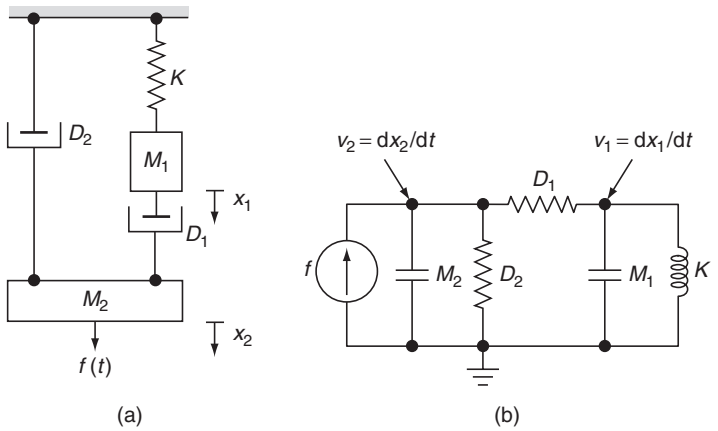


Figure 8.8 A two co-ordinate mechanical system: (a) Mechanical system; (b) electrical analog.

A mechanical system with two-co-ordinate movement and its equivalent electrical circuit are shown in Figure 8.8, where K represents a spring and D_1 and D_2 represent the dashpots.

In this case the equations, written directly in terms of the Laplace transform variables, are obtained by applying KCL at each of the two ungrounded nodes.

$$(s^2M_2 + sD_2 + sD_1)X_2(s) - sD_1X_1(s) = F(s)$$

$$sD_1X_2(s) + (s^2M_1 + sD_1 + K)X_1(s) = 0 \tag{8.23}$$

The convenience of writing Eq. (8.23) in terms of node voltages is evident. If one needs to obtain the transfer function relating $X_2(s)$ to $F(s)$, solve Eq. (8.23) for $X_2(s)$. Hence

$$\frac{X_2(s)}{F(s)} = \frac{s^2M_1 + sD_1 + K}{(s^2M_1 + sD_1 + K)(s^2M_2 + sD_2 + sD_1) - s^2D_1^2} \tag{8.24}$$

Modeling an Armature-Controlled DC Motor

Figure 8.9 shows the schematic diagram of an armature-controlled DC servomotor. It will be assumed that the field current is maintained constant and a voltage $v(t)$ is applied to the armature, which has a resistance R_a and negligible inductance. The effect of the application of the input voltage $v(t)$ will cause the armature to rotate. Applying the KVL to the electrical side of the system, one obtains

$$v(t) = R_a i_a(t) + L_a \frac{di_a}{dt} + v_b(t) \tag{8.25}$$

where $v_b(t) = \text{back emf} = K \cdot \omega$ and K is the motor back emf constant which is dependent on the motor parameters and the field applied to the field winding. However, for a constant field, K is a constant.

The torque developed by the motor is given by

$$T_d = K \cdot i_a$$

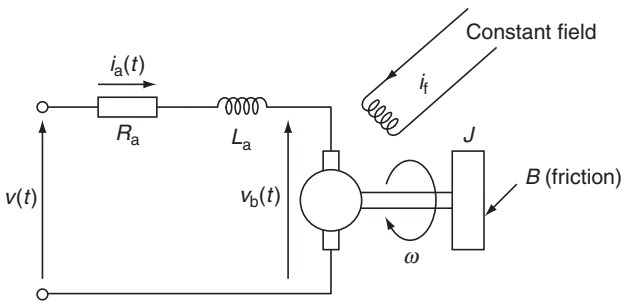


Figure 8.9 | Armature-controlled DC servomotor.

This developed torque has to work against the acceleration torque needed to overcome the inertia J and the friction B. Further if the shaft of the motor is connected to any other system, an additional load T_L is reflected onto the shaft of the motor. The developed torque should work against this also. Therefore, the developed torque and the angular velocity ω are related as

$$T_d = K \cdot i_a = J\frac{d\omega}{dt} + B\omega + T_L \tag{8.26}$$

Equations (8.25) and (8.26) are the two first-order linear differential equations that define the behavior of the constant field DC motor system. Taking i_a and ω as the state variables, the state equation is obtained by re-writing Eqs. (8.25) and (8.26) as

$$\begin{bmatrix} di_a/dt \\ d\omega/dt \end{bmatrix} = \begin{bmatrix} -R/L_a & -K/L_a \\ K/J & -B/J \end{bmatrix} \cdot \begin{bmatrix} i_a \\ \omega \end{bmatrix} + \begin{bmatrix} 1/L_a & 0 \\ 0 & -1/J \end{bmatrix} \cdot \begin{bmatrix} v \\ T_L \end{bmatrix} \tag{8.27}$$

Equation (8.27) is the state equation for the DC motor system wherein v and T_L are the external excitations that are applied to the system.

To obtain the model in the transfer function form, one can use Eq. (8.27) or one can take the Laplace transform of Eqs. (8.25) and (8.26). This gives

$$V(s) = R_a I_a(s) + L_a s I_a(s) + K\omega(s) \tag{8.28}$$

$$KI_a(s) = Js\omega(s) + B\omega(s) + T_L(s) \tag{8.29}$$

Transfer functions are defined for one specific output variable and one specific input excitation making all other excitations as zero. Therefore to obtain the transfer function for $\omega(s)/V(s)$, the load torque input is taken as zero. Substituting Eq. (8.29) into Eq. (8.28) and simplifying one obtains

$$G(s) = \frac{\omega(s)}{V(s)} = \frac{K/L_a J}{s^2 + [(B/J) + (R_a/L_a)]s + [(R_a B/L_a J) + (K^2/L_a J)]} \tag{8.30}$$

Equations (8.28) and (8.29) can be visualized in the form of a block diagram as shown in Figure 8.10. This block diagram has been obtained using the Laplace transforms of Eqs. (8.28) and (8.29).

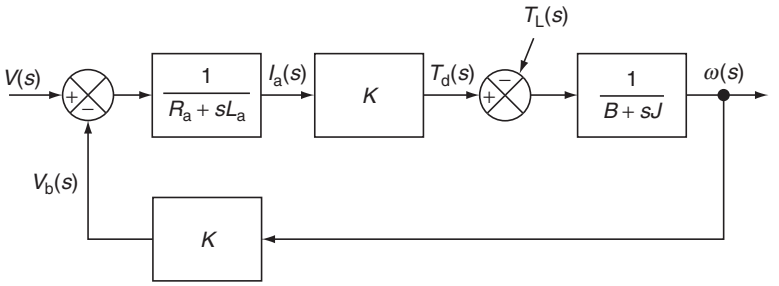


Figure 8.10 Block diagram showing the various relationships in the armature-controlled DC motor.

This simple example illustrates the use of the block diagram representation. It can be simplified through certain rules of block diagram algebra. These will be discussed in the following section. It must be emphasized that the model obtained for the DC motor is based on several simplifying assumptions. The armature reaction in the motor and the voltage drops in the brushes have been neglected. In addition, it is assumed that the frictional torque is linear and directly proportional to the angular velocity. This is true only over some operating range.

Poles and Zeros

The transfer functions are represented in the form of a numerator polynomial and a denominator polynomial in terms of the Laplace variable. In general, the transfer function for any system is of the form

$$\frac{Y(s)}{U(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0} \tag{8.31a}$$

The order of the denominator polynomial "n" indicates the order of the system. It is in general greater than or equal to the order of the numerator polynomial "m" for a causal system. However, in the case of predictive systems wherein there is a differentiator element, "m" will be greater than "n". In Eq. (8.31a), the numerator and denominator polynomials can be represented in the factored form as

$$\frac{Y(s)}{U(s)} = \frac{(s + n_1) \cdot (s + n_2) \cdots (s + n_m)}{(s + d_1) \cdot (s + d_2) \cdots (s + d_n)} \tag{8.31b}$$

From Eqs. (8.31a) and (8.31b), it is evident that there are "m" factors in the numerator and "n" factors in the denominator. The Laplace variable "s" can take on values that are both real and complex. At the values of $s = -n_1, -n_2, \ldots, -n_m$, the transfer function is zero. The values $-n_1, -n_2, \ldots, -n_m$ are called the *zeros* of the system. Likewise, when the values of $s = -d_1, -d_2, \ldots, -d_n$, the transfer function is infinity. The values $-d_1, -d_2, \ldots, -d_n$ are called the *poles* of the system.

8.5 Block Diagrams

In the analysis of control systems it is very convenient to obtain the block diagrams of different components and their interconnections. If the various components are non-interacting (i.e., there is no "loading" effect of one component on another), it is possible to obtain the overall transfer function of the system through a

suitable combination of the transfer functions of the component blocks utilizing some basic rules of block diagram transformations to reduce the original diagram. It should be understood that absence of loading effect means that each block of the block diagram has infinite input impedance and zero output impedance. This is a rather stringent constraint in block diagram algebra which is not true in practical systems. Most practical systems have finite input and output impedances and therefore the transfer functions resulting from block diagram simplifications must be used with care and awareness.

For example, consider the cascade blocks shown in Figure 8.11(a). Here,

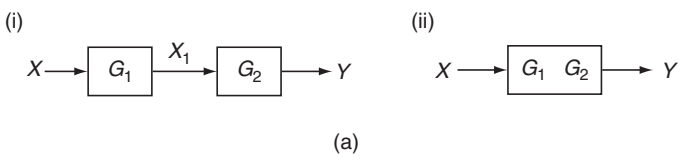
$$X_2 = G_1 X_1 \tag{8.32}$$

$$Y = G_2 X_2 \tag{8.33}$$

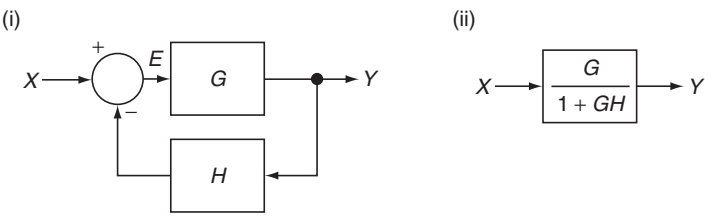
Substituting Eq. (8.32) in Eq. (8.33) one obtains

$$Y = G_1 G_2 X_1 \tag{8.34}$$

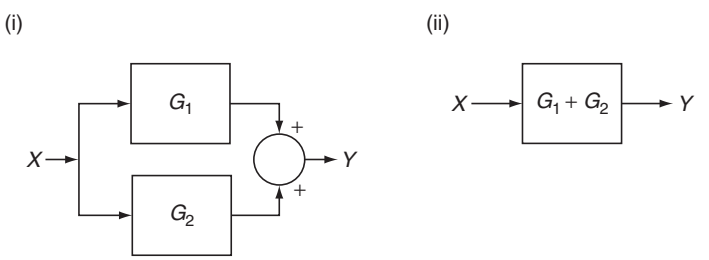
This is shown in the equivalent block diagram alongside in Figure 8.11(a).



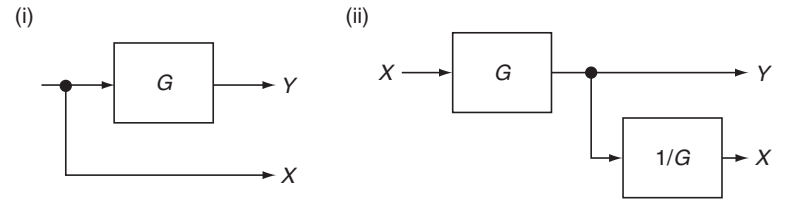
(a)



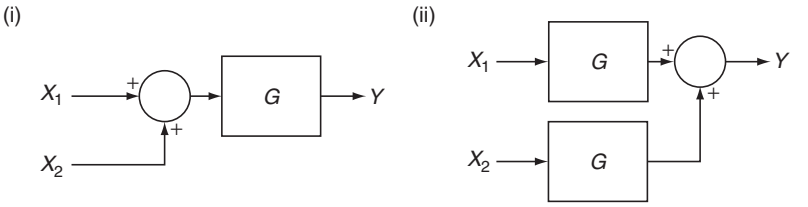
(b)



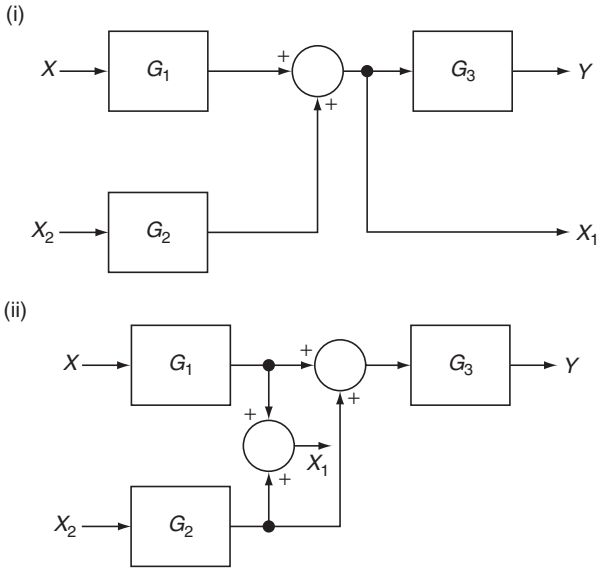
(c)


(i)
G
Y
X
(ii)
X
G
Y
1/G
X

(d)


(i)
X1
+
G
Y
+
X2
(ii)
X1
G
+
Y
+
X2
G

(e)


(i)
X
G1
+
G3
Y
+
X2
G2
X1
(ii)
X
G1
+
G3
Y
+
+
X1
+
X2
G2

(f)

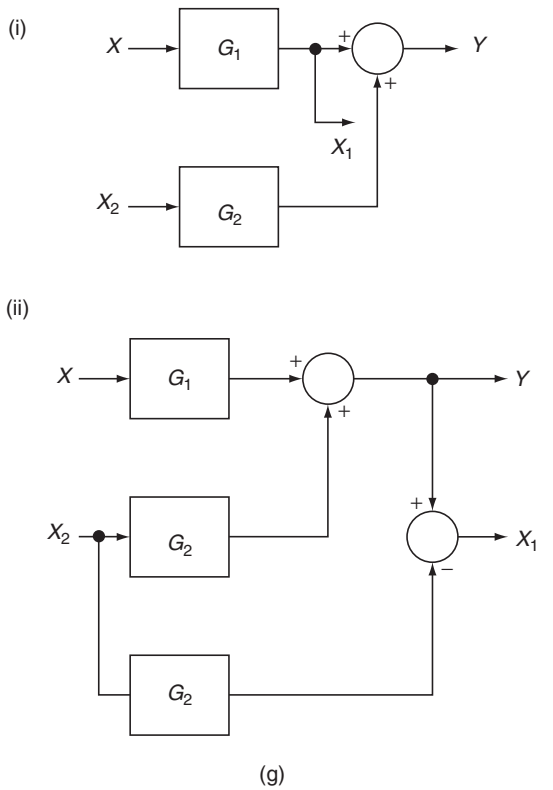


Figure 8.11 Block diagram reduction rules: (a) Combining blocks in cascade; (b) elimination of feedback loop; (c) combining blocks in parallel; (d) moving a pick-off point behind a block; (e) moving a summing point behind a block; (f) moving a pick-off point ahead of a summing point; (g) moving a pick-off point behind a summing point.

Similarly, for the feedback loop in Figure 8.11(b), one obtains

$$Y = GE \tag{8.35}$$

and

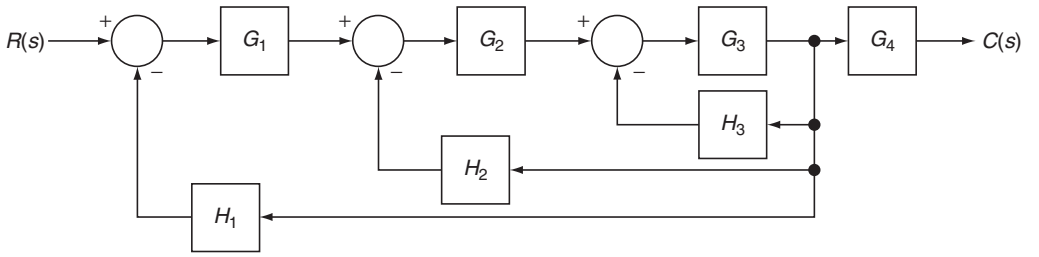
$$E = X - HY \tag{8.36}$$

Eliminating E from the two equations by substituting Eq. (8.36) into Eq. (8.35) and after simplification, one obtains

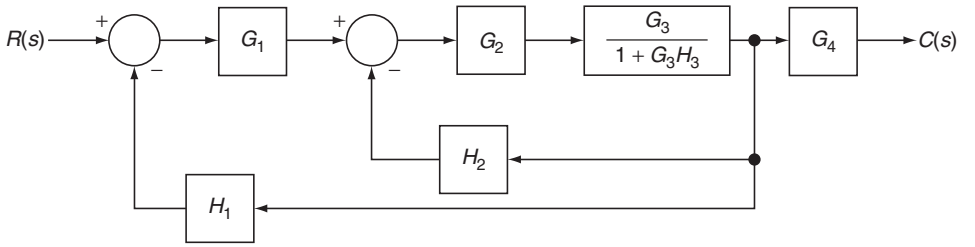
$$Y = \frac{G}{1+GH} X \tag{8.37}$$

The proofs for the remaining transformations can be obtained easily in a similar manner and hence are left to the reader as an exercise. One may apply the block diagram results given in Eqs. (8.34) and (8.37) for the DC motor block diagram of Figure 8.10 to obtain the DC motor transfer function as given in Eq. (8.30).

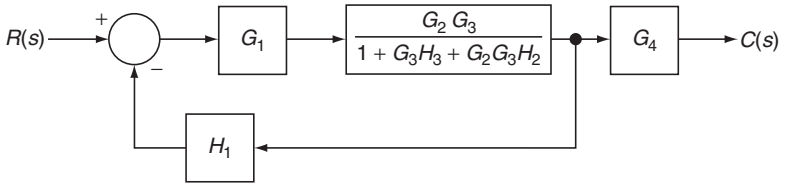
As an example, consider the block diagram shown in Figure 8.12(a). Based on the reduction rules as given in Figure 8.11, one can try to simplify the block diagram. In this case, the first step is to replace the innermost loop by its equivalent transfer function. The resulting block diagram is shown in Figure 8.12(b). Proceeding in the same manner, one can again remove the next inner loop after combining the two blocks in cascade. The resulting block diagram is shown in Figure 8.12(c).



(a)



(b)



(c)

Figure 8.12 (a) An example block diagram; (b) simplified block diagram; (c) further simplification of the block diagram.

Finally, reducing the last feedback loop, one gets the overall transfer function as

$$\frac{C(s)}{R(s)} = \frac{(G_1G_2G_3G_4)/(1+G_3H_3+G_2G_3H_2)}{1+[(G_1G_2G_3)/(1+G_3H_3+G_2G_3H_2)]H_1}$$
$$= \frac{G_1G_2G_3G_4}{1+G_3H_3+G_2G_3H_2+G_1G_2G_3H_1} \tag{8.38}$$

Mason's Rule

The overall transfer function can be obtained directly from the block diagram by using Manson's rule, given as follows:

$$G(s) = \frac{C(s)}{R(s)} = \frac{\sum_k T_k(s)\Delta_k(s)}{\Delta(s)} \tag{8.39}$$

where $T_k(s)$ is the transfer function of the kth forward path from the input to the output; $\Delta(s)$ is the determinant of the block diagram; $\Delta_k(s)$ all terms in $\Delta(s)$ that do not have elements or paths common with an element or path in $T_k(s)$. Mathematically $\Delta(s)$ is represented as

$\Delta s = 1 -$ sum of all individual loop transfer functions
$+$ sum of the products of the transfer functions of all possible sets of two non-touching loops
$-$ sum of the productsof the transfer functions of all possible sets of three non-touching loops
$+ \&$

Note that two loops are said to be non-touching if they do not have a common branch or node. Also, a loop is a closed path in the direction of the arrows that does not retrace itself.

As an example, consider the block diagram shown in Figure 8.12(a). Here, there is only one forward path from $R(s)$ to $C(s)$, so that $k = 1$, and three loops with transmittances $-G_3H_3$, $-G_2G_3H_2$ and $-G_1G_2G_3H_1$. Also, all the loops are touching each other. Thus

$$T_1 = G_1G_2G_3G_4$$
$$\Delta = 1 - (-G_3H_3 - G_2G_3H_2 - G_1G_2G_3H_1)$$
$$= 1 - (-G_3H_3 - G_2G_3H_2 - G_1G_2G_3H_1)$$

Hence

$$G(s) = \frac{T_1}{\Delta} = \frac{G_1G_2G_3G_4}{1 + G_3H_3 + G_2G_3H_2 + G_1G_2G_3H_1}$$

which is as given in Eq. (8.38).

EXAMPLE 8.2

Consider the block diagram shown in Figure 8.13. The transfer function $C(s)/R(s)$ can be determined by block diagram simplification. This is done most conveniently by replacing the inner loop by its equivalent and then adding the two parallel branches to obtain a single loop diagram. One can apply Mason's rule and the result can be verified by the reader by using the procedure suggested above.

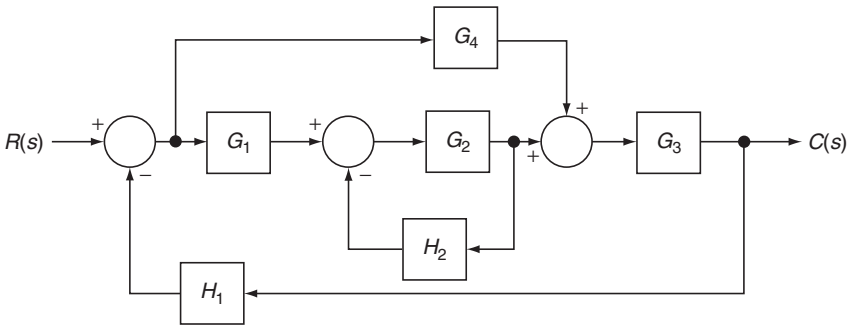


Figure 8.13 An example of block diagram.

Here, there are two forward paths from $R(s)$ to $C(s)$ and

$$T_1(s) = G_1G_2G_3$$

$$T_2(s) = G_4G_3$$

Further, there are three loops with transmittances $-G_2H_2$, $-G_1G_2G_3H_1$ and $-G_4G_3H_1$. Furthermore, the first and the last loops are non-touching. This gives

$$\Delta(s) = 1 + (G_2H_2 + G_1G_2G_3H_1 + G_4G_3H_1)$$

$$\Delta_1(s) = 1$$

$$\Delta_2(s) = 1 + G_2H_2$$

Hence

$$\frac{C(s)}{R(s)} = \frac{G_1G_2G_3 + G_4G_3(1+G_2H_2)}{1+G_2H_2+G_1G_2G_3H_1+G_4G_3H_1+G_2H_2G_4G_3H_1}$$

8.6 Lagrange Method

The 18th-century French mathematician Lagrange developed a method to obtain the equations governing the motion of complex mechanical systems. The differential equations resulting from this method are known as Lagrange's equations. They are basically derived from Newton's laws of motion. Even though the method was initially applied to mechanical systems, it is generic enough to be applied to other energy domains as it deal with scalar quantities like potential and kinetic energy rather than vector quantities like forces and torques.

The fundamental concept of Lagrange method is the representation of the system by a set of generalized co-ordinates, one for each independent degree of freedom of the system. After having defined the generalized co-ordinates, the kinetic energy E_K is expressed in terms of these co-ordinates and their derivatives. The potential energy E_P is expressed in terms of the generalized co-ordinates. If the generalized co-ordinates are e_i where $i = 1,2, \ldots, r$, then the Lagrangian function L is given by

$$L = E_K(e_1,\ldots,e_r,\dot{e}_1,\ldots,\dot{e}_r) - E_P(e_1,\ldots,e_r) \tag{8.40}$$

Finally, the desired equations of motion called the Lagrange's equation are derived using the Lagrangian function as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{e}}\right) - \frac{\partial L}{\partial e_i} = F_i \quad i = 1,2, \ldots, r \tag{8.41}$$

where F_i's denote the generalized forces that are external to the system. They are obtained from the algebraic sum of the external forces and non-energy-storing/dissipative forces acting on the ith co-ordinate.

In the case of mechanical systems, the generalized co-ordinates are the independent linear or the angular displacements of the different masses of the systems. The constitutive relations are used to express all the forces on the elements in terms of the displacement and its derivatives. Though the Lagrange method was developed and used for mechanical systems to model the dynamics of motion of bodies, the same concept can be extended to electrical networks to model the dynamics of motion of charges. In electrical networks, the charges can be considered to form the generalized co-ordinates. A systematic procedure to select the generalized co-ordinates for electrical systems is to take the independent loop currents to be the time derivatives of the generalized

charge co-ordinates. Using the constitutive relations for the various elements, all the voltages across all circuit elements are expressed in terms of the charges and their derivatives. The Lagrange method can be understood by the following two examples, one for the mechanical system and the other for the electrical network.

The kinetic energy E_K of the system is the sum of the kinetic energy of each mass. The cart is confined to move only in the horizontal direction and therefore its kinetic energy is given as

$$E_{K1} = \frac{1}{2} M\dot{x}^2$$

The pendulum rod rotates about the hinge. Therefore, the pendulum mass has one component of motion along the horizontal and another along the vertical. The kinetic energy of the pendulum mass is given by

$$E_{K2} = \frac{1}{2} m(\dot{x}_2^2 + \dot{h}^2)$$

EXAMPLE 8.3

Inverted Pendulum

The cart with an inverted pendulum, shown in Figure 8.14, is excited by a controlled input force *F*. Assume that the pendulum does not move more than a few degrees away from the vertical. The problem is to determine the dynamic equations of motion for this system in the state space form.

It is seen that the motion of the system is defined by the displacement *x* of the cart from a reference point and the angle θ that the pendulum rod makes with respect to the vertical. Therefore, the system has only two degrees of freedom and the dynamics are expressed in terms of the corresponding generalized co-ordinates which in this case is (x, θ).

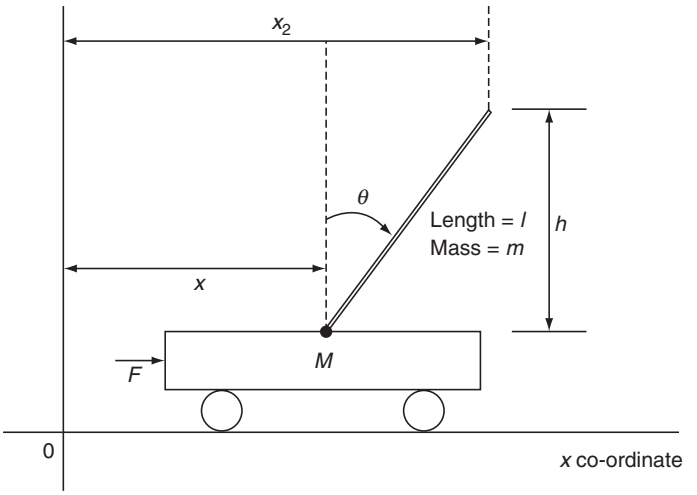


Figure 8.14 Inverted pendulum on a moving cart. Here *M* is the mass of the cart; *m* the mass of the pendulum; *l* the length of the pendulum; *F* the force applied to the cart; *x* the cart position; θ the pendulum angle from the vertical.

where

$$x_2 = x + l\sin\theta$$
$$\dot{x}_2 = \dot{x} + l\dot{\theta}\cos\theta$$
$$h = l\cos\theta$$
$$\dot{h} = -l\dot{\theta}\sin\theta$$

Now the total kinetic energy of the system is given by

$$E_K = E_{K1} + E_{K2} = \frac{1}{2}M\dot{x}^2 + \frac{1}{2}m(\dot{x}^2 + 2\dot{x}\dot{\theta}l\cos\theta + l^2\dot{\theta}^2)$$

In the vertical direction, if the cart is considered as the reference, then the potential energy is that stored in the pendulum. It is given as

$$E_P = mgh = mgl\cos\theta$$

Now the Lagrangian function is given as

$$L = E_K - E_P = \frac{1}{2}M\dot{x}^2 + \frac{1}{2}m(\dot{x}^2 + 2\dot{x}\dot{\theta}l\cos\theta + l^2\dot{\theta}^2) - mgl\cos\theta$$

The generalized co-ordinates are (x, θ). The Lagrange's equations for this system are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = F \tag{8.42}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0 \tag{8.43}$$

Equations (8.42) and (8.43) are the Lagrange's equations for this system. The partial derivatives of the Lagrangian with respect to the generalized co-ordinates are given as

$$\frac{\partial L}{\partial \dot{x}} = (M+m)\dot{x} + ml\dot{\theta}\cos\theta$$
$$\frac{\partial L}{\partial x} = 0$$
$$\frac{\partial L}{\partial \dot{\theta}} = ml\dot{x}\cos\theta + ml^2\dot{\theta}$$
$$\frac{\partial L}{\partial \theta} = mgl\sin\theta$$

Substituting the above partial derivatives in the Lagrange's equations [Eqs. (8.42) and (8.43)], one obtains

$$\begin{aligned}(M+m)\ddot{x} + ml\ddot{\theta}\cos\theta - ml\dot{\theta}^2\sin\theta &= F\\ ml\ddot{x}\cos\theta - ml\dot{x}\dot{\theta}\sin\theta + ml^2\ddot{\theta} - mgl\sin\theta &= 0\end{aligned} \tag{8.44}$$

Equation (8.44) gives the exact equations of motion of the inverted pendulum mounted on a moving cart. Evidently, the equations are non-linear differential equations due to the presence of $\sin\theta, \cos\theta, \dot{\theta}^2, \dot{x}\dot{\theta}$

terms. Under equilibrium condition, the pendulum rod is in the vertical position. If one considers that the deviation in θ is kept small, then in the neighborhood about the vertical operating point, the system may be considered as linear. Therefore, in the very small neighborhood about the vertical operating point,

$$\cos\theta \approx 1$$
$$\sin\theta \approx \theta$$

It may be assumed that the derivatives $\dot{x}$ and $\dot{\theta}$ are kept small and therefore the quadratic terms $\dot{\theta}^2$ and $\dot{x}\dot{\theta}$ will also be negligible. Applying these linearizing assumptions to the non-linear equations of motion given in Eq. (8.44), one obtains the linearized dynamic model given as

$$\begin{aligned}(M+m)\ddot{x} + ml\ddot{\theta} &= F \\ m\ddot{x} + ml\ddot{\theta} - mg\theta &= 0\end{aligned} \tag{8.45}$$

There are two second-order linear differential equations. This means that the overall order of the system is four. This implies that there must be four state variables to represent the state space of the system. The state vector containing the four state variables is

$$\mathbf{x} = \begin{bmatrix} x \\ \theta \\ \dot{x} \\ \dot{\theta} \end{bmatrix}$$

The four first-order linear differential equations representing the dynamic behavior of the system in terms of the state variables and the external excitation are

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \dot{x}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \dot{\theta}$$

From Eq. (8.45), the other two differential equations are

$$\frac{\mathrm{d}\dot{x}}{\mathrm{d}t} = \frac{-mg}{M}\theta + \frac{1}{M}F \tag{8.46}$$

$$\frac{\mathrm{d}\dot{\theta}}{\mathrm{d}t} = \frac{M+m}{Ml}g\theta + \frac{-1}{Ml}F \tag{8.47}$$

From the four first-order linear differential equations listed above, the state space representation of the system in the standard form is given as

$$\begin{bmatrix} \dot{x} \\ \dot{\theta} \\ \ddot{x} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -mg/M & 0 & 0 \\ 0 & (M+m)g/Ml & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ \theta \\ \dot{x} \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1/M \\ -1/Ml \end{bmatrix} \cdot F \tag{8.48}$$

$$\theta = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ \theta \\ \dot{x} \\ \dot{\theta} \end{bmatrix} \tag{8.49}$$

Equations (8.48) and (8.49) give the state and the output equations for the inverted pendulum system. To obtain the model in the transfer function form, one can use Eq. (8.20) or take the Laplace transforms of Eq. (8.45). On simplification and accounting for the two integrator for the state variables x and θ, one obtains the fourth-order transfer function as

$$G(s) = \frac{\theta(s)}{F(s)} = \frac{-1/Ml}{s^2 - [(M+m)g/Ml]} \tag{8.50}$$

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{Ms^2} \frac{s^2 - (Mg/Ml)}{s^2 - [(M+m)g/Ml]} \tag{8.51}$$

EXAMPLE 8.4

Electrical Network

Consider the electrical network shown in Figure 8.15. Each independent loop current will be chosen as independent charge variables q_1 and q_2. These two charge variables form the generalized co-ordinates for the Lagrange method. The voltages across all the elements of the circuit are expressible in terms of the charge variables and their derivatives. Thus,

$v_R = R \cdot (\dot{q}_1 - \dot{q}_2)$ is the voltage across the resistor R

$v_{L1} = L_1 \cdot \ddot{q}_1$ is the voltage across the inductor L_1

$v_{L2} = L_2 \cdot \ddot{q}_2$ is the voltage across the inductor L_2

$v_{C1} = q_1 / C_1$ is the voltage across the capacitor C_1

$v_{C2} = q_2 / C_2$ is the voltage across the capacitor C_2

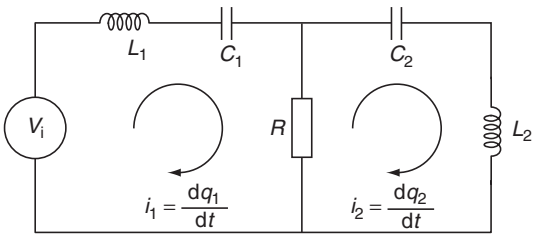


Figure 8.15 | Example of electrical network.

The Lagrangian function L is given as

$$L = E_K - E_P = \frac{1}{2} \cdot L_1 \cdot \dot{q}_1^2 + \frac{1}{2} \cdot L_2 \cdot \dot{q}_2^2 - \frac{q_1^2}{2C_1} - \frac{q_2^2}{2C_2} \tag{8.52}$$

The Lagrange's equations for this system with $[q_1, q_2]$ as the generalized co-ordinates are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_1}\right)-\frac{\partial L}{\partial q_1}=\sum_{\text{loop1}} V_{\text{non-storing-devices}}=V_i-R\cdot(\dot{q}_1-\dot{q}_2) \tag{8.53}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_2}\right)-\frac{\partial L}{\partial q_2}=\sum_{\text{loop2}} V_{\text{non-storing-devices}}=R\cdot(\dot{q}_1-\dot{q}_2) \tag{8.54}$$

Equations (8.53) and (8.54) are the Lagrange's equations for this system. The partial derivatives of the Lagrangian with respect to the generalized co-ordinates are given as

$$\frac{\partial L}{\partial \dot{q}_1}=L_1\cdot\dot{q}_1$$

$$\frac{\partial L}{\partial q_1}=\frac{-q_1}{C_1}$$

$$\frac{\partial L}{\partial \dot{q}_2}=L_2\cdot\dot{q}_2$$

$$\frac{\partial L}{\partial q_2}=\frac{-q_2}{C_2}$$

Substituting these partial derivatives in the Lagrange's equations [Eqs. (8.53) and (8.54), one obtains

$$L_1\cdot\ddot{q}_1+\frac{q_1}{C_1}=V_i-R\cdot(\dot{q}_1-\dot{q}_2) \tag{8.55}$$

$$L_2\cdot\ddot{q}_2+\frac{q_2}{C_2}=R\cdot(\dot{q}_1-\dot{q}_2) \tag{8.56}$$

Equations (8.55) and (8.56) give the dynamic model of the electrical network of Figure 8.15. These two equations can be re-arranged and split into four linear first-order differential equations using $\ddot{q}_1$, $\dot{q}_1$, $\ddot{q}_2$ and $\dot{q}_2$ as state variables to obtain the standard state equation representation.

8.7 Circuit Averaging

The Lagrange method uses scalar variables like the kinetic energy and the potential energy of the components in the system. This does not directly give information on the power flow direction. Another problem with the Lagrange method is that it cannot easily model switched power systems and the dynamics related to such systems like the switched-mode DC–DC converters. Circuit averaging method is a useful modeling approach especially for switched power electric circuits. The circuit averaging method is explained in the following steps.

Step 1 (Large-Signal Model): Identify the active circuit for the different position of the switch. From the equivalent active circuit for each switch position, write down the dynamic circuit equations using the Kirchhoff's voltage and current laws. The state variables are the inductor currents and capacitor voltages. The state equations of the active circuit for the various switch positions are written down. For a converter circuit using a single-pole double-throw switch, there are two switch positions and therefore, two operative circuit modes. The two state equations are of the form

$$\dot{x} = A_1 x + B_1 u \tag{8.57}$$

$$y = C_1 x + D_1 u$$

$$\dot{x} = A_2 x + B_2 u \tag{8.58}$$

$$y = C_2 x + D_2 u$$

Equations (8.57) and (8.58) are the state equations for the two operative circuit modes of a converter with a two-state switch.

Step 2 (Average Large-Signal Model): The state equations represent the dynamic model of the active circuit resulting from the specific switch position for a specified interval of time. The above representation is in the standard state space format for each of the intervals. If it is a two-state switch then there are two sets of state equations. The converter alternates between the two switched states at high frequency. It is required to represent the converter through a single equivalent dynamic representation valid for both states of the switch.

Consider a converter system with two switch states. The pole of the switch is at throw 1 for a period of time dT_s and at throw 2 for a period of time $(1 - d)T_s$. If one considers the variation of the state variables over a switching period, then

$$x = \dot{x}_{\text{avg}} T_s = \dot{x}_{dT_S} \cdot dT_s + \dot{x}_{(1-d)T_S} \cdot (1-d)T_s$$

where $\dot{x}_{\text{avg}}$ is the average rate of change of state variables over a switching period. The above description is valid if $\dot{x}_{dT_S}$ and $\dot{x}_{(1-d)T_S}$ are constant during the dT_s and $(1 - d)T_s$ duration, respectively. This will be a valid assumption if the switching period is small compared to the natural time constants of the respective circuits. Then the averaged state variables are obtained from the state equations [Eqs. (8.57) and (8.58)] as

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{8.59}$$

where

$$A = A_1 d + A_2 (1-d)$$

$$B = B_1 d + B_2 (1-d)$$

$$C = C_1 d + C_2 (1-d)$$

$$D = D_1 d + D_2 (1-d)$$

Equation (8.59) represents the equivalent state equation of the converter. Since the averaging process has been done over a switching period, the equivalent model is valid for time durations much larger compared to the switching period.

Step 3 (Steady-State Model): The steady-state solution is obtained by equating the rate of change of the state variables to zero. Under steady-state or equilibrium conditions,

$$\begin{aligned} \dot{x} &= 0 \\ x &= X \\ u &= U \\ y &= Y \end{aligned} \tag{8.60}$$

For the steady-state conditions, Eq. (8.59) can be written as

$$\begin{aligned} 0 &= AX + BU \\ Y &= CX + DU \end{aligned} \tag{8.61}$$

Step 4 (Small Signal Model): The averaged large-signal model of the converter given by Eq. (8.59) is linear but not time invariant. This is because the characteristic matrix A and the input matrix B contain the duty ratio variable d that is time-varying. Therefore it is necessary to model the system in the neighborhood of the operating point to analyze the system and synthesize controllers for the system. Such a model is called the small-signal model. From Eq. (8.59), the averaged dynamic equations are

$$\begin{aligned} \dot{x} &= [A_1 d + A_2(1-d)]x + [B_1 d + B_2(1-d)]u \\ y &= [C_1 d + C_2(1-d)]x + [D_1 d + D_2(1-d)]u \end{aligned} \tag{8.62}$$

All the variables are considered to have small variations in the neighborhood of the steady-state or equilibrium state operating point. Thus,

$$d = D + \hat{d}; \qquad \frac{\hat{d}}{D} \ll 1$$

$$u = U + \hat{u}; \qquad \frac{\hat{u}}{U} \ll 1$$

The small-signal variations in duty ratio d and input u result in perturbations in x and y about the operating points. Thus,

$$x = X + \hat{x}; \qquad \frac{\hat{x}}{X} \ll 1$$

$$y = Y + \hat{y}; \qquad \frac{\hat{y}}{Y} \ll 1$$

Therefore Eq. (8.62) can be written as

$$\begin{aligned} (\dot{X} + \dot{\hat{x}}) &= [A_1(D+\hat{d}) + A_2(1-D-\hat{d})](X+\hat{x}) + [B_1(D+\hat{d}) + B_2(1-D-\hat{d})](U+\hat{u}) \\ (Y+\hat{y}) &= [C_1(D+\hat{d}) + C_2(1-D-\hat{d})](X+\hat{x}) + [D_1(D+\hat{d}) + D_2(1-D-\hat{d})](U+\hat{u}) \end{aligned} \tag{8.63}$$

The above equations can be expanded and separated into steady-state terms, linear small-signal terms and non-linear terms. The non-linear terms contain the second- and higher order perturbation products which can be neglected.

Terms containing $\hat{x} \cdot \hat{d}$ and $\hat{d} \cdot \hat{u}$ can be neglected and the steady-state terms

$$AX + BU = 0$$

Thus, the linear small-signal terms are

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + B\hat{u} + [(A_1 - A_2)X + (B_1 - B_2)U]\hat{d} \\ \hat{y} &= C\hat{x} + D\hat{u} + [(C_1 - C_2)X + (D_1 - D_2)U]\hat{d} \end{aligned}$$

The perturbation in duty ratio $\hat{d}$ is now considered as an input for the small-signal model. Thus, the resultant small-signal model is given as

$$\begin{aligned}\dot{\hat{x}} &= A\hat{x} + B\hat{u}_n \\ \hat{y} &= C\hat{x} + D\hat{u}_n\end{aligned} \tag{8.63}$$

where

$$A = A_1 D + A_2(1-D) \tag{8.63a}$$

$$B = \begin{bmatrix} B_1 D + B_2(1-D) & (A_1 - A_2)X + (B_1 - B_2)U \end{bmatrix} \tag{8.63b}$$

$$\hat{u}_n = \begin{bmatrix} \hat{u} \\ \hat{d} \end{bmatrix} \tag{8.63c}$$

$$C = C_1 D + C_2(1-D) \tag{8.63d}$$

$$D = \begin{bmatrix} D_1 D + D_2(1-D) & (C_1 - C_2)X + (D_1 - D_2)U \end{bmatrix} \tag{8.63e}$$

$$X = -A^{-1}BU \tag{8.63f}$$

EXAMPLE 8.5

Buck Converter

Consider the buck converter as discussed in Chapter 5. The active circuits for the two states of the single-pole double-throw (SPDT) switch are shown in Figure 8.16. The operation and steady-state analysis of the converter is as discussed in Chapter 5. Referring to Figure 8.16, as there are two energy-storage devices, there will be need for two state variables. The dynamic inductor element stores energy by virtue of the flow of current through it and therefore, the inductor current i_L can be considered as a state variable. The dynamic capacitor element stores energy by virtue of the voltage across it and therefore, the capacitor voltage v_C can also be considered as a state variable. Thus the state vector is

$$x = \begin{bmatrix} i_L \\ v_C \end{bmatrix} \tag{8.64}$$

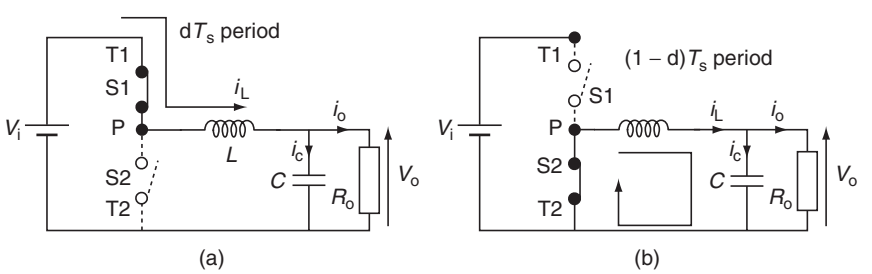


Figure 8.16 The operative circuit during: (a) Period dT_s; (b) period $(1-d)T_s$.

Step 1 (Large-Signal Model):
During dT_s Period: Referring to Figure 8.16(a), the state equations during this period are obtained by applying the KVL for the voltage across the inductor and the KCL for the current through the capacitor. Thus,

$$\begin{aligned}\frac{\mathrm{d}i_\mathrm{L}}{\mathrm{d}t} &= \frac{v_\mathrm{i} - v_\mathrm{C}}{L} \\ \frac{\mathrm{d}v_\mathrm{C}}{\mathrm{d}t} &= \frac{i_\mathrm{L} - i_\mathrm{o}}{C} = \frac{i_\mathrm{L} - v_\mathrm{C}/R_\mathrm{o}}{C}\end{aligned} \tag{8.65}$$

The state equation is

$$\begin{bmatrix}\frac{\mathrm{d}i_\mathrm{L}}{\mathrm{d}t} \\ \frac{\mathrm{d}v_\mathrm{C}}{\mathrm{d}t}\end{bmatrix} = \begin{bmatrix}0 & \frac{-1}{L} \\ \frac{1}{C} & \frac{-1}{R_\mathrm{o}C}\end{bmatrix} \cdot \begin{bmatrix}i_\mathrm{L} \\ v_\mathrm{C}\end{bmatrix} + \begin{bmatrix}\frac{1}{L} \\ 0\end{bmatrix} \cdot \begin{bmatrix}v_\mathrm{i}\end{bmatrix} \tag{8.66a}$$

and the output equation is

$$v_\mathrm{o} = \begin{bmatrix}0 & 1\end{bmatrix} \cdot \begin{bmatrix}i_\mathrm{L} \\ v_\mathrm{C}\end{bmatrix} + \begin{bmatrix}0\end{bmatrix} \cdot \begin{bmatrix}v_\mathrm{i}\end{bmatrix} \tag{8.66b}$$

Equations (8.66a) and (8.66b) are of the form

$$\dot{x} = A_1 \cdot x + B_1 \cdot u$$
$$y = C_1 x + D_1 u$$

During $(1 - d)T_s$ Period: Referring to Figure 8.16(b), the state equations during this period are obtained by applying the KVL for the voltage across the inductor and the KCL for the current through the capacitor. Thus,

$$\begin{aligned}\frac{\mathrm{d}i_\mathrm{L}}{\mathrm{d}t} &= \frac{0 - v_\mathrm{C}}{L} \\ \frac{\mathrm{d}v_\mathrm{C}}{\mathrm{d}t} &= \frac{i_\mathrm{L} - i_\mathrm{o}}{C} = \frac{i_\mathrm{L} - v_\mathrm{C}/R_\mathrm{o}}{C}\end{aligned} \tag{8.67}$$

The state equation is

$$\begin{bmatrix}\frac{\mathrm{d}i_\mathrm{L}}{\mathrm{d}t} \\ \frac{\mathrm{d}v_\mathrm{C}}{\mathrm{d}t}\end{bmatrix} = \begin{bmatrix}0 & \frac{-1}{L} \\ \frac{1}{C} & \frac{-1}{R_\mathrm{o}C}\end{bmatrix} \cdot \begin{bmatrix}i_\mathrm{L} \\ v_\mathrm{C}\end{bmatrix} + \begin{bmatrix}0 \\ 0\end{bmatrix} \cdot \begin{bmatrix}v_\mathrm{i}\end{bmatrix} \tag{8.68a}$$

and the output equation is

$$v_\mathrm{o} = \begin{bmatrix}0 & 1\end{bmatrix} \cdot \begin{bmatrix}i_\mathrm{L} \\ v_\mathrm{C}\end{bmatrix} + \begin{bmatrix}0\end{bmatrix} \cdot \begin{bmatrix}v_\mathrm{i}\end{bmatrix} \tag{8.68b}$$

Equations (8.68a) and (8.68b) are of the form

$$\dot{x} = A_2 \cdot x + B_2 \cdot u$$
$$y = C_2 x + D_2 u$$

Step 2 (Averaged Large-Signal Model): The averaged large-signal model is obtained from Eq. (8.59). T he state equations [Eqs. (8.66) and (8.68)] for the two operative intervals are combined to obtain the averaged large-signal model. Thus,

$$\begin{bmatrix} \frac{\mathrm{d}i_\mathrm{L}}{\mathrm{d}t} \\ \frac{\mathrm{d}v_\mathrm{C}}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & \frac{-1}{L} \\ \frac{1}{C} & \frac{-1}{R_\mathrm{o}C} \end{bmatrix} \cdot \begin{bmatrix} i_\mathrm{L} \\ v_\mathrm{C} \end{bmatrix} + \begin{bmatrix} \frac{d}{L} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} v_\mathrm{i} \end{bmatrix} \tag{8.69a}$$

$$v_\mathrm{o} = \begin{bmatrix} 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} i_\mathrm{L} \\ v_\mathrm{C} \end{bmatrix} \tag{8.69b}$$

Equations (8.69a) and (8.69b) give the state equation and the output equation of the averaged large-signal model.

Step 3 (Steady-State Model): The steady-state model is obtained based on Eqs. (8.60) and (8.61). Thus the steady-state model is given as

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{-1}{L} \\ \frac{1}{C} & \frac{-1}{R_\mathrm{o}C} \end{bmatrix} \cdot \begin{bmatrix} I_\mathrm{L} \\ V_\mathrm{C} \end{bmatrix} + \begin{bmatrix} \frac{D}{L} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} V_\mathrm{i} \end{bmatrix} \tag{8.70a}$$

$$V_\mathrm{o} = \begin{bmatrix} 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} I_\mathrm{L} \\ V_\mathrm{C} \end{bmatrix} \tag{8.70b}$$

Equations (8.70a) and (8.70b) are the state and output equations of the steady-state model. The steady-state input–output relationship that is discussed in Chapter 5 is obtainable from the above equations. From Eq. (8.70) it can be seen that

$$V_\mathrm{o} = DV_\mathrm{i}$$

Step 4 (Small Signal Model): The small-signal model is obtained from Eqs. (8.63a)–(8.63f). The small-signal model is given as

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + B\hat{u}_n \\ \hat{y} &= C\hat{x} + D\hat{u}_n \end{aligned} \tag{8.71}$$

where

$$A = \begin{bmatrix} 0 & \frac{-1}{L} \\ \frac{1}{C} & \frac{-1}{R_\mathrm{o}C} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{D}{L} & \frac{D}{L}V_\mathrm{i} \\ 0 & 0 \end{bmatrix}$$

$$\hat{u}_n = \begin{bmatrix} \hat{v}_i \\ \hat{d} \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

Equation (8.71) is the small-signal model of the buck converter where $\hat{d}$ is another input to the system. Normally, $\hat{d}$ is the control input to the buck converter. The transfer function $V_o(s)/d(s)$ can be obtained from the small-signal model using Eq. (8.20).

8.8 Bond Graphs

Bond graph is a graphical tool for modeling physical systems. In 1959, Prof. H. M. Paynter gave the revolutionary idea of portraying systems in terms of power bonds, connecting the elements of the physical system to the so-called junction structures. The power exchange picture of a system is called bond graph/energy bond graph/power bond graph that can be both power- and information-oriented. "Bond graphs" is a system modeling technique. It has been used extensively in modeling interconnected interacting physical systems. Bond graphs not only allow the modeling of systems for analysis and simulation, but they are also a powerful tool for automatic computer programming. Bond graphs were invented by Henry Paynter due to the inherent drawbacks of block diagrams for servocontrols and simulation problems. Bond graphs consider both energy and signal exchanges between components of a system. For systems where power and efficiency play a major role, bond graph modeling method is very convenient.

By this approach, a physical system can be represented by symbols and lines, identifying the power flow paths. The lumped parameter elements of resistance, capacitance and inductance are interconnected in an energy conserving way by bonds and junctions resulting in a network structure. From the pictorial representation of the bond graph, the derivation of system equations is so systematic that it can be algorithmized.

Bond graphs display both energy and signal exchanges between components or elements in installations and systems by simple lines and symbols. They bridge the gap between control engineering and the parts of engineering science where power and efficiency have greater importance including energy conversion.

Bond graphs are a representation of components, machines and systems on paper using conventional signs and symbols. Bond graphs are set up causally (without such a distinction between cause and effect) and transformed subsequently into a causal diagram by a systematic choice of the causalities.

Block Diagram Drawbacks

1. A return action between components must always be shown by a separate feedback connection, while in reality the components have only one connection.
2. Block diagrams for a given physical system can have a radically different appearance because of the different possible causalities.
3. The loading effect of subsequent blocks is not reflected backwards.

Equivalent Circuits Drawbacks

1. Due to presence of fictitious elements the equivalent circuit loses its similarity with the real layout.
2. With each symbol, it must be indicated separately, whether it is a real component or an idealized element.
3. Some of the representation may be in a domain in which the actual system does not exist. For example, the thermal model of the system being represented as an electrical equivalent.

Table 8.2 | Effort and flow variables for systems in few energy domains

Systems	Effort (*e*)	Flow (*f*)
Mechanical	Force (F)	Velocity (v)
	Torque (τ)	Angular velocity (ω)
Electrical	Voltage (V)	Current (i)
Hydraulic	Pressure (P)	Volume flow rate (dQ/dt)
Thermal	Temperature (T)	Entropy change rate (dS/dt)
Chemical	Chemical potential (μ)	Mole flow rate (dN/dt)
	Enthalpy (h)	Mass flow rate (dm/dt)
Magnetic	Magneto-motive force (e_m)	Rate of flux change ($d\Phi/dt$)

Bond graphs avoid the above drawbacks and can represent interdisciplinary systems with one set of symbols for all disciplines. Bond graphs are based on the splitting of systems into separate components that exchange energy or power through identifiable connections or ports. They are called bonds in analogy to the energy exchange between atoms in chemical bonds.

In this technique, power flow is represented by a *half bond*. Every bond is associated with two variables, effort and flow and the causality indication. The various components in a system linked using the bonds form the *power bond graph*. Any system can be modeled using the finite set of elements that are one-port, two-port or multi-port depending on the number of ports for exchanging energy.

1. **One-port:** These include sources (effort and flow), passive elements (dissipation, kinetic and potential storage).
2. **Two-port:** These include transformer, modulated transformers, gyrators and modulated gyrators.
3. **Multi-port:** These include 0-junction and 1-junction.

The energy flux or power in a bond is always the product of two variables – a potential variable called "effort" and a "flow" variable or current variable, simply referred as flow. Table 8.2 gives the effort and flow of various domains.

Standard Elements

The standard elements of bond graphs are classified according to the number of bonds as one-ports, two-ports and multi-ports. This classification originates from electric circuit theory where each port or bond represents two wire terminals or connections. A one-port has two connections, a two-port four connections, etc.

One-Ports

One-ports are elements exchanging energy with the system through one bond only. They include resistance elements (called *R*-elements), inertia elements (called *L*-elements) and capacity elements (called *C*-elements). In addition there are the voltage sources and the current sources.

Each element has two possible governing equations based on the causality. In the particular application, the causality will indicate the governing equation for the bond. The causality is shown as a vertical bar either at the beginning or at the end of the bond. Table 8.3 gives the one-port elements.

Table 8.3 | One-ports

Bond graph element	Governing equations	Remarks
Sources		
S_e ——▷|	E	Effort source
S_f |—▷	F	Flow source
mS_e ——▷|	E	Modulated effort source
mS_f |—▷	F	Modulated flow source
Sinks – Dissipative and Storage Elements		
——▷| R	$f = \frac{e}{R}$	Dissipative effort causal
|—▷ R	$e = f \cdot R$	Dissipative flow causal
——▷| L	$f = \frac{1}{L}\int e \cdot dt$	Kinetic energy storage with effort causal. Note the integral cause. The effect which is *flow* is a STATE variable.
|—▷ L	$e = L \cdot \frac{df}{dt}$	Flow causal. Differential causal.
——▷| C	$f = C \cdot \frac{de}{dt}$	Effort causal. Differential causal.
|—▷ C	$e = \frac{1}{C}\int i \cdot dt$	Potential energy storage with effort causal. The effect which is *effort* is a state variable.

Two-Ports

Two-ports have two bonds for energy exchange with the system. They are energy conserving in the sense that the product of input flow and effort is at all times equal to the product of output flow and effort. Broadly there are two types of two-ports based on the type of energy variable linking between the input and output ports: (a) transformers and (b) gyrators.

In transformers, the cause-and-effect variables are both either efforts or flows. This means that there is only input effort–output effort link or input flow–output flow linkage. Transformers, although they are not normally distinguished in bond graph, are of the following two different kinds.

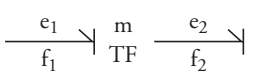
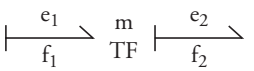
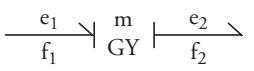
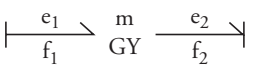
1. **Impedance Transformers:** Here the input and output variables are of the same energy domain or class. They change the impedances, that is, the ratio of effort to flow in the bonds. Example: Electrical transformers or gear reducers.

2. **Class Transformers:** Here input and output variables belong to different energy domains or classes or disciplines. Example: Hydraulic cylinders and pumps connecting hydraulic with mechanical variables because the impedance here belongs to different classes (have different physical dimensions), it is not possible to compare them.

Class transformers are important because they allow an efficient representation of interdisciplinary engineering systems by bond graphs. It should be mentioned again that both class and impedance transformers conserve the power at all times.

In the case of gyrators, the effort of one-port is linked to the flow of the other port. Thus one has cross-linkage of variables in the case of gyrators like input effort–output flow or input flow–output effort linkage. Table 8.4 gives the list of two-ports elements.

Table 8.4 | Two-ports

Bond graph element	Governing equations	Remarks
	$e_2 = m \cdot e_1$ $f_1 = m \cdot f_2$	Transformer effort causal. e_1 is the cause and f_1 is the effect.
	$f_2 = \frac{1}{m} \cdot f_1$ $e_1 = \frac{1}{m} \cdot e_2$	Transformer flow causal. f_1 is the cause and e_1 is the effect.
	Same as transformer with effort causal, but m is modulated (varying).	Modulated transformer effort causal
	Same as transformer with flow causal, but m is modulated (varying).	Modulated transformer flow causal
	$f_2 = \frac{1}{m} \cdot e_1$ $f_1 = \frac{1}{m} \cdot e_2$	Gyrator effort causal. e_1 is the cause and f_1 is the effect.
	$e_2 = m \cdot f_1$ $e_1 = m \cdot f_2$	Gyrator flow causal. f_1 is the cause and e_1 is the effect.
	Same as gyrator with effort causal, but m is modulated (varying).	Modulated gyrator effort causal.
	Same as gyrator with flow causal, but m is modulated (varying).	Modulated gyrator flow causal.

Multi-Ports (Junctions)

The elements of bond graphs are connected by either of the following two junctions:

1. The **0-junction** where the efforts (voltages) on all bonds are same and the flows (currents) are additive or, more precisely, add up to zero. This junction corresponds to the parallel circuit in electronics.
2. The **1-junction** where the efforts (voltages) on all bonds add up to zero and the flows (currents) on all bonds are same. It corresponds to the series circuit in electronics.

Junctions are power conserving at each instant and the power transports of all bonds add up to zero at all times. Table 8.5 describes the 0-junction and the 1-junction.

Rules for the Selection of Causality

Causality is not a natural fact but selected by the engineer to facilitate computation. The choice is free in principle but subject to the following limitations:

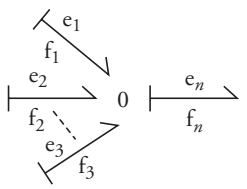
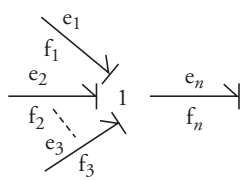
1. Sources always impose an effort or a flow to the system. Consequently the causality must be chosen first.
2. With one-port elements, the integral causal is the preferred causality by nature. These also decide the state variables of the system. The dynamic elements are then chosen for integral causality.
3. The junction rules for 0-junction and 1-junction should then be applied.
4. The dissipative elements do not have any preferred causality. Causality of these bonds are chosen to have consistent causality for the whole system.

Steps in Obtaining the System Model

System analysis using bond graph is very organized and systematic. Given any system the following simple steps should be executed to arrive at the model of the physical system:

1. Construct the bond graph for the given system.
2. Assign causality using following order:

Table 8.5 | Multi-ports

Bond graph element	Governing equations	Remarks
	$e_1 = e_2 = \ldots = e_n$ $\sum_{i=1}^{n} f_i = 0$	**0-junction:** There can be only one bond with a causal bar at the 0-junction. This means that only one bond can decide the junction effort. The algebraic sum of all the associated bond flows should be zero. This is called the FLOW LAW. Example: Parallel circuits in electrical domain (KCL).
	$\sum_{i=1}^{n} e_i = 0$ $f_1 = f_2 = \ldots = f_n$	**1-junction:** There can be only one bond without a causal bar at the 1-junction. This means that only one bond can decide the junction flow. The algebraic sum of all the associated blond efforts should be zero. This is called the EFFORT LAW. Example: Series circuits in electrical domain (KVL).

- Sources.
- Dynamic elements (Integral equation is given preference over differential equation.).
- Junction rule to be applied.
- Dissipative element.

3. **State Equation Extraction:** The order of the system as well as the total number of equations needed to describe the system depends upon the number of storage elements. Write down the state equations considering the causality and the effort-flow relationship for the various elements.

Bond Graph Construction

The above steps in obtaining the state space model will be illustrated with a number of examples. However, one must first learn to construct the bond graph for a given system. There are two ways in which the bond graph may be constructed: (a) by inspection approach and (b) by algorithmic approach. The former is very intuitive and lends itself very well to modeling complex physical systems depending on the amount of knowledge and experience the modeler has with the particular system. The latter method is algorithmic and lends itself to computer-aided model extraction. This section will discuss both the approaches.

Consider the following *RLC* circuit as shown in Figure 8.17. To obtain the state space model for the *RLC* system, one has to first construct the bond graph. Both the methods of bond graph construction will be illustrated for the circuit of Figure 8.17.

Inspection Approach

The *RLC* circuit of Figure 8.17 can be re-written as shown in Figure 8.18(a). The circuit consists of the voltage source e_1, the resistance R_1, the inductance L and block B1. All these components share the same current; therefore, the corresponding bond graph for these elements is as shown in Figure 8.18(a). The main property of the 1-junction is that all bonds connected to it share a common flow. The flow in this case is the current. Therefore, components having same current must be connected to a 1-junction. On studying the bond graph of Figure 8.18(a), it is observed that there are four bonds at the 1-junction:

- An energy bond for the effort source Se_1.
- An energy bond for energy flowing into the resistance R_1 and dissipating.
- An energy bond for the energy flowing into the inductance L and stored.
- An energy bond for the energy flowing into the block B1.
- The arrows indicate the general direction of the flow of energy. It should be noted that the instantaneous power can also flow in a direction opposite to that indicated in the bond arrows in which case the corresponding power variable (either the voltage/effort or the current/flow) is negative. However, the bond arrows should, in general, be placed to indicate the energy flow.

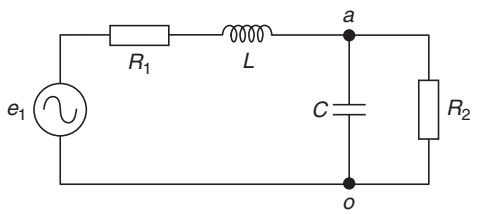


Figure 8.17 | *RLC* circuit.

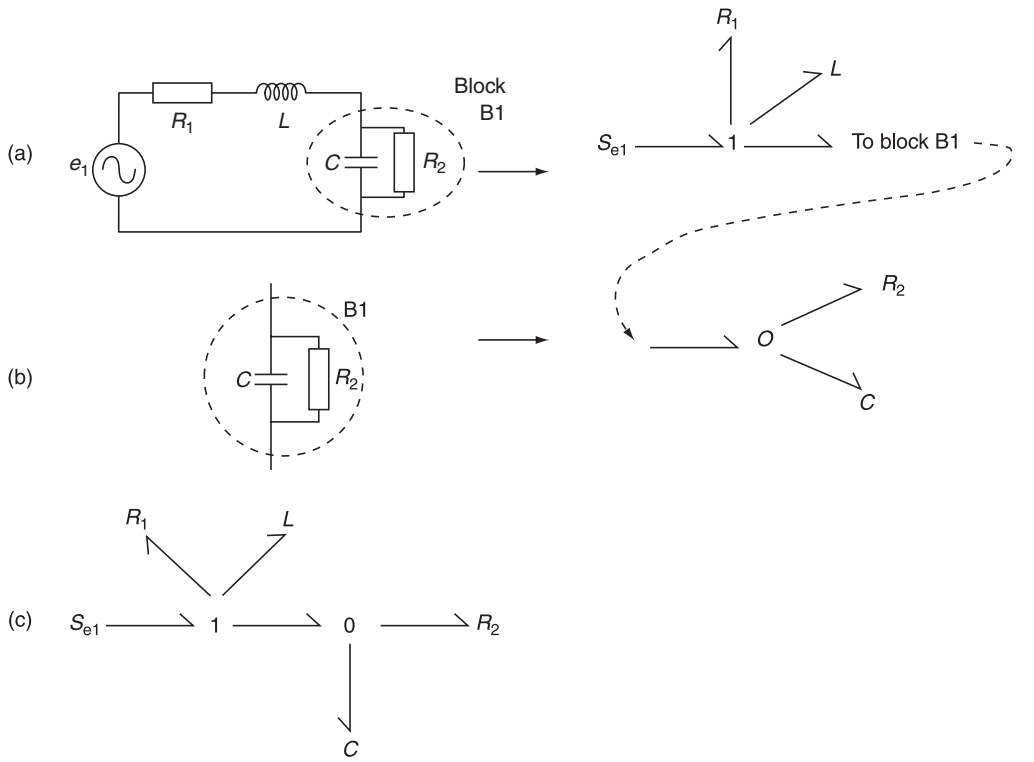


Figure 8.18 (a) *RLC* circuit and bond graph; (b) block B1 and bond graph; (c) overall bond graph.

Figure 8.18(b) shows block B1 which has C and R_2 in parallel. Both the components of block B1 share a common voltage/effort. Therefore, one should use a 0-junction to connect the two components because in the case of a 0-junction all the bonds connected to it share a common effort. Referring to the bond graph of block B1 in Figure 8.18(b), the 0-junction consists of

1. the energy bond from the 1-junction that indicates the flow of energy to block B1.
2. an energy bond indicating the flow of energy into resistance R_2.
3. another energy bond to indicate the flow of energy into capacitance C for storage.

Figure 8.18(c) shows the completed bond graph by concatenating the bond graphs of Figures 8.18(a) and 8.18(b). The bond graph of Figure 8.18(c) can also be arrived at by the algorithmic approach. This approach will now be explained.

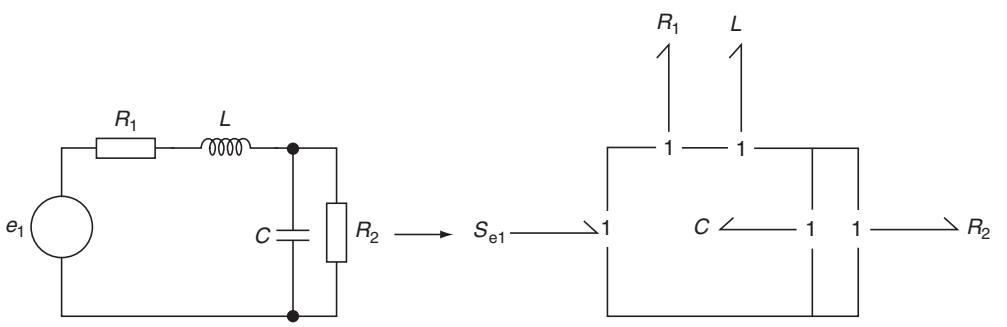
Algorithmic Approach

The bond graph may also be constructed in an algorithmic way that will prove useful for computer-aided generation of bond graphs for physical systems. However, this approach lacks the intuition that one gains from the method of inspection. The sequential steps involved in this method are as follows:

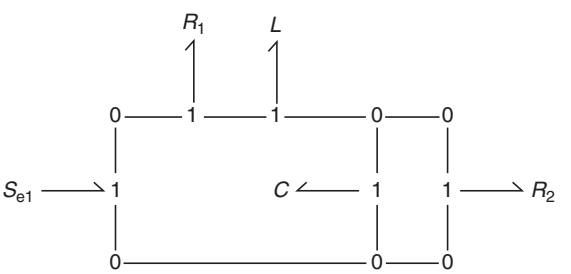
1. Replace every component by a 1-junction and connect the appropriate one-port bond to the junction.
2. Introduce 0-junctions for all nodal points in the circuit.

3. Combine all contiguous 1-junctions into a single 1-junction.
4. Combine all contiguous 0-junctions into a single 0-junction.
5. Remove any junction that has less than three bonds connected to it and again apply Steps (3) and (4) if necessary.
6. Identify a ground node (any of the 0-junction nodes).
7. Remove all bonds connected to ground node 0-junction including the junction.
8. Apply Steps (3), (4) and/or (5) again.

Applying the above steps to the *RLC* circuit of Figure 8.17, the sequence of bond graph construction is shown in Figure 8.19. Figure 8.19(a) shows the application of Step (1) to the RLC circuit. Here every component is replaced with a 1-junction. A bond is used to connect the component to the 1-junction. The bond direction is decided on the energy flow direction, that is, energy flows out of sources and into the sinks like *R*, *L* and *C*. Figure 8.19(b) shows the application of Step (2). Here a 0-junction is introduced at every node as shown in Figure 8.19(b). Figure 8.19(c) shows the application of Steps (3) and (4). Contiguous 1-junctions are combined into a single 1-junction. Likewise contiguous 0-junctions are combined into a single 0-junction. As 1-junctions and 0-junctions are multi-ports, 1- and 0-junctions with less than three energy bonds connected to them can be replaced with a single bond. This is illustrated in Figure 8.19(d). After this, once again the process of reducing contiguous junctions is performed.



(a) Step 1



(b) Step 2

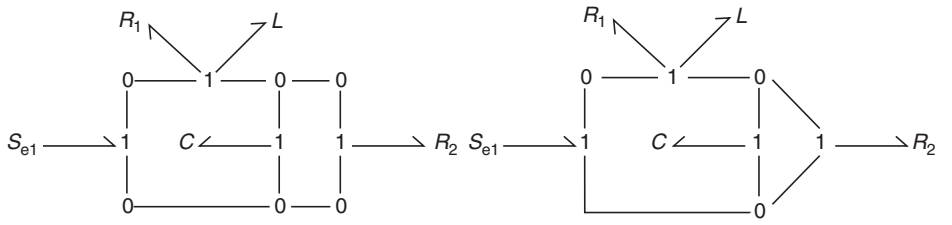


Figure 8.19 (a) Applying Step (1) to the *RLC* circuit; (b) applying Step (2); (c) bond reduction by applying Steps (3) and (4); (d) applying Step (5); (e) applying Steps (6) and (7); (f) applying Step (8).

Now a reference 0-junction has to be chosen. Any 0-junction can be chosen as the reference 0-junction. In Figure 8.19(e), one such 0-junction is considered as the reference. After choosing the reference 0-junction, remove that junction along with all the energy bonds connected to that 0-junction. Finally reducing the bond graph by applying Steps (3), (4) and/or (5), the final bond graph is obtained which is identical to the bond graph obtained by the inspection approach as given in Figure 8.18(c).

Causality Assignment

After the bond graph has been constructed, the next step is to assign causality to each bond. The assignment of causality gives an insight into the internal working of the system at each bond level. As there are two power variables associated with each bond, one will act as the cause variable and the other will be the effect variable. The causality assignment will define the cause-and-effect relationship between the effort and the flow variables across the entire system. The general rules for causality assignment have been mentioned previously. The causality assignment sequence is as follows:

1. Sources always impose an effort or a flow to the system. Therefore, the causality of the effort and/or the flow sources must be chosen first. The causality is assigned as given in the section "Rules for Selection of Causality".
2. The causality of bonds connected to dynamic elements (L and C) is chosen next. Referring to the section "Steps in Obtaining the System Model", the causality assignment for the dynamic elements is chosen such that the governing equation is an integration rather than differentiation. This is because nature prefers continuum. This means that for an inductive element, the effort causal assignment is preferred and for a capacitive element, the flow causal assignment is preferred.
3. Next, the 0-junction and the 1-junction rules are applied to propagate the causality to other bonds. It should be noted that at the 0-junction only one bond can decide the effort at the junction, that is, only one bond can have a causal bar at the junction end and all other bonds should have the causal bar away from the junction end. In the case of the 1-junction, only one bond can decide the flow at the junction, that is, only one bond can have a causal bar away from the junction end and all the other bonds should have the causal bar towards the junction end.
4. The causality of the two ports is chosen in a manner such that there is no conflict with the above three rules.
5. The dissipative elements do not have any preferred causality. Causality of these bonds is chosen in the end to have consistent causality for the whole system.

The bond graph which also shows the causality assignment is called the augmented bond graph. The above causality assignment rules will be applied to the bond graph of the *RLC* circuit given in Figure 8.19(f). The causality assignment sequence is shown in Figure 8.20. Figure 8.20(a) begins the causality assignment sequence by defining the source causality. The effort source causality is assigned such that it has the causal bar at the arrow end.

Next the causality of the dynamic elements is considered. Without loss of generality one can start with any of the dynamic elements. The causality for the L-bond is assigned such that the governing equation has an integration. Thus the causal bar for the L-bond is at the arrow end which indicates effort causal, that is, for the L-bond, the effort or voltage across the inductor is the independent variable or the cause variable and the flow or the current through the inductor is the dependent variable or the effect variable. The flow effect is related to the effort cause by means of an integral equation as given in Table 8.3. Likewise the C-bond is also assigned the causality such that the governing equation has an integration. Referring to the discussion on the one-ports and Table 8.3, the causal bar is at the non-arrow end. This means that the flow or the current is the cause and the effort or voltage is the effect which is obtained as the output of the integral equation. Figure 8.20(b) shows the causality assignment for the bonds connected to dynamic elements.

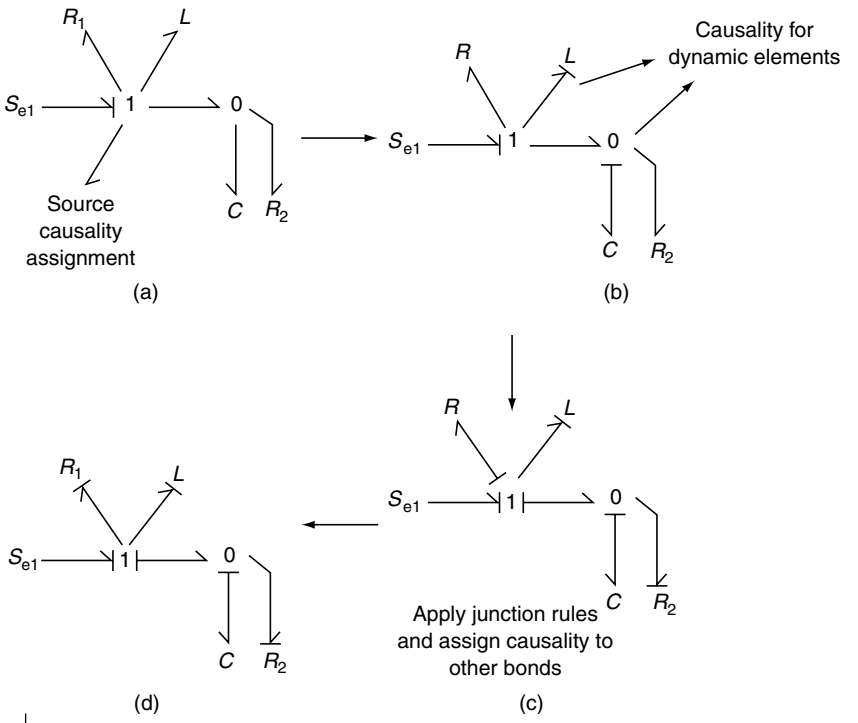


Figure 8.20 (a) Assign causality to source bonds; (b) assign causality to dynamic element bonds; (c) assign causality to other bonds based on junction rules; (d) final augmented bond graph of *RLC* circuit.

This is followed by assigning causality to bonds connected to the 1- and 0-junction by applying the junction rules. At the 1-junction all the bonds except one should have the causal bar at the junction end of the bond. As the *L*-bond already has the causal bar at the arrow end, all the other bonds connected to the 1-junction should have the causal bar at the junction end. For the case of the 0-junction only one bond should have a causal bar at the junction end. As the C-bond already has the causal bar at the junction end, all other bonds connected to the 0-junction should have the bonds away from the 0-junction end. Without loss of generality, one can assign causality to the bonds using the junction rules to as many bonds as possible such that there are no causal conflicts. Figure 8.20(c) shows the causality assignment by applying the junction rules.

In the end, only the bonds connected to dissipative elements like resistances will be left. The dissipative elements do not have any preferred causality as their governing equations are only algebraic. Thus the remaining bonds that have not been assigned any causality will be those connected to dissipative elements. Here any causality can be assigned such that the junction rule is not violated. The complete augmented bond graph is shown in Figure 8.20(d).

State Equation Extraction

The final part of the modeling process is the extraction of the state equation from the augmented bond graph. The state equation extraction sequence is very simple and straightforward. With some amount of practice, one can write down the state equation of any system just by inspection of the augmented bond graph for the system. The state equation extraction sequence is as follows:

1. Define the state variables and the input variables.
2. Number the junctions that are connected to dynamic elements.
3. Write the effort and flow variables of all the bonds connected to these numbered junctions in terms of the state variables and input variables only.
4. Apply junction equations to obtain the first-order linear differential equations.
5. There will be as many equations as there are dynamic elements with integral causality. These equations are merged together to be represented in matrix form called the state equation.

Figure 8.21 illustrates the application of the state equation extraction algorithm. First the input and state variables are identified. Figure 8.21(a) shows the input and state variables for the *RLC* circuit. Here V_{in} is the input variable. Current through the inductor i_L and the voltage across the capacitor v_C are the state variables. There are two junctions in the system bond graph of the *RLC* circuit. Both these junctions are associated with dynamic elements and therefore both these junctions are uniquely numbered. Junction 1 is associated with the inductive dynamic element and junction 2 is associated with the capacitive dynamic element.

Next, the efforts and the flows of all the bonds connected to these numbered junctions will have to be identified in terms of the input and state variables. This is shown in Figure 8.21(b). Junction 1 is a 1-junction and therefore all the bonds connected to it will share the same flow, that is, i_L. Junction 2 is a 0-junction and therefore all the bonds connected to it will share the same effort, that is, v_C. The effort of the R_1-bond is obtained from its governing equation as $i_L R_1$. Likewise, the flow of R_2-bond is obtained from its governing equation as v_C / R_2. All the bond efforts and flows for the two numbered junctions are shown in Figure 8.21(b).

In the next step the linear differential equations can be read out by inspection by applying the junction rule which states that for a 1-junction, the algebraic sum of all efforts of bonds connected to the 1-junction is zero and for the 0-junction, the algebraic sum of all flows of bonds connected to the 0-junction is zero. It should be noted that bond arrows give an indication of the energy flow and therefore efforts and flows of bonds where energy enters the junction are considered positive and the efforts and flows of bonds wherein the energy leaves the junction are considered negative.

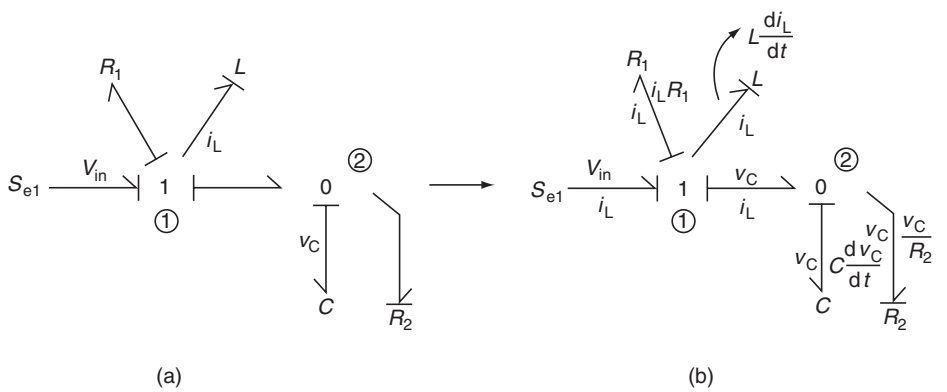


Figure 8.21 (a) Defining input and state variables; (b) efforts and flows of bonds connected to junctions that are associated with dynamic elements.

Considering junction 1:
As it is a 1-junction, one should apply the rule that the algebraic sum of all the efforts of bonds connected to this junction is zero. Therefore

$$V_{in} - L\frac{di_L}{dt} - i_L R_1 - v_C = 0 \tag{8.72}$$

The state derivative of Eq. (8.72) can be written as a function of input and state variables by re-arranging Eq. (8.72). Thus,

$$\frac{di_L}{dt} = -\frac{R_1}{L} i_L - \frac{1}{L} v_C + \frac{1}{L} V_{in} \tag{8.73}$$

Considering junction 2:
As this is a 0-junction, one should apply the rule that the algebraic sum of the flows of all the bonds connected to this junction is zero. Therefore,

$$i_L - C\frac{dv_C}{dt} - \frac{v_C}{R_2} = 0 \tag{8.74}$$

The state derivative of Eq. (8.74) can be written as a function of input and state variables by re-arranging Eq. (8.74). Thus,

$$\frac{dv_C}{dt} = \frac{1}{C} i_L - \frac{1}{R_2 C} v_C + 0 \cdot V_{in} \tag{8.75}$$

Equations (8.73) and (8.75) give the dynamic behavior of the system. As there are two dynamic elements in the *RLC* circuit, only two first-order linear differential equations are sufficient to fully describe its dynamic behavior. These two equations can be combined into matrix representation to obtain the state equation as

$$\begin{bmatrix} \frac{di_L}{dt} \\ \frac{dv_C}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R_1}{L} & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{R_2 C} \end{bmatrix} \cdot \begin{bmatrix} i_L \\ v_C \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} \cdot V_{in} \tag{8.76}$$

Equation (8.76) gives the state equation of the *RLC* circuit of Figure 8.17. The process of obtaining the state equation for the *RLC* circuit has been explained in detail in this section. The same procedure can be applied to any physical system to obtain the state space model.

DC Motor Examples

Separately Excited DC Motor Figure 8.22(a) shows the schematic of a separately excited DC motor. Using the method discussed above, the bond graph model with causality assigned is shown in Figure 8.22(b). Every junction has a subscripted number that indicates the junction number. This is used in order to identify the junction during equation formation. The state equation is obtained from the bond graph model by the method discussed in this section. Observe that the state equation obtained is same as that derived for the armature-controlled DC motor of Figure 8.9.

$$\begin{bmatrix} \frac{di_a}{dt} \\ \frac{d\omega}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R_a}{L_a} & -\frac{k}{L_a} \\ \frac{k}{J} & -\frac{B}{J} \end{bmatrix} \begin{bmatrix} i_a \\ \omega \end{bmatrix} + \begin{bmatrix} \frac{1}{L_a} & 0 \\ 0 & -\frac{1}{J} \end{bmatrix} \begin{bmatrix} v \\ T_L \end{bmatrix} \tag{8.77}$$

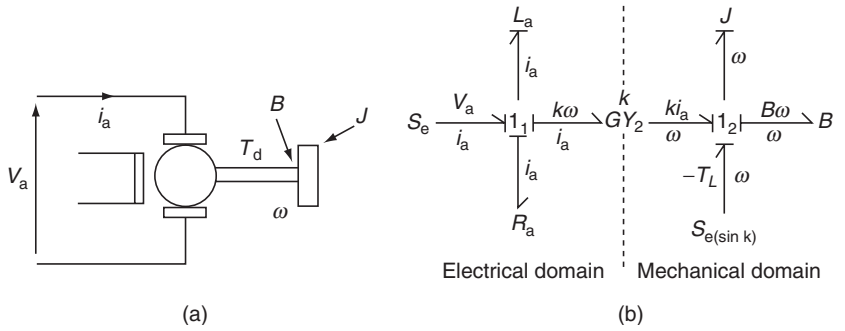


Figure 8.22 | (a) Separately excited DC motor; (b) bond graph with assigned causality.

DC Motor Excited by Armature Supply The schematic of a DC motor that is excited by the same armature supply that also supplies power to the armature is shown in Figure 8.23(a). The bond graph is developed along similar lines discussed in this section and is shown in Figure 8.23(b). Observe that there are three state variables as there are three dynamic elements: (a) armature inductance, (b) field inductance and (c) mechanical inertia. Further from the bond graph given in Figure 8.23(b), there is state variable multiplication in the mGY component that multiplies the armature current and the field current to produce the torque. The back effect of the gyrator that produces the back emf is also a product of the two state variables: angular speed and field current. This makes the model non-linear. The set of non-linear large-signal differential equations are given in Eqs. (8.78a), (8.78b), (8.78c). One can obtain the linearized small-signal model by considering perturbations in the neighborhood of the operating point as discussed in the section "Differential Equations and Linearization".

$$\frac{\mathrm{d}i_a}{\mathrm{d}t} = -\frac{R_a}{L_a} i_a - \frac{ki_f\omega}{L_a} + \frac{v}{L_a} \tag{8.78a}$$

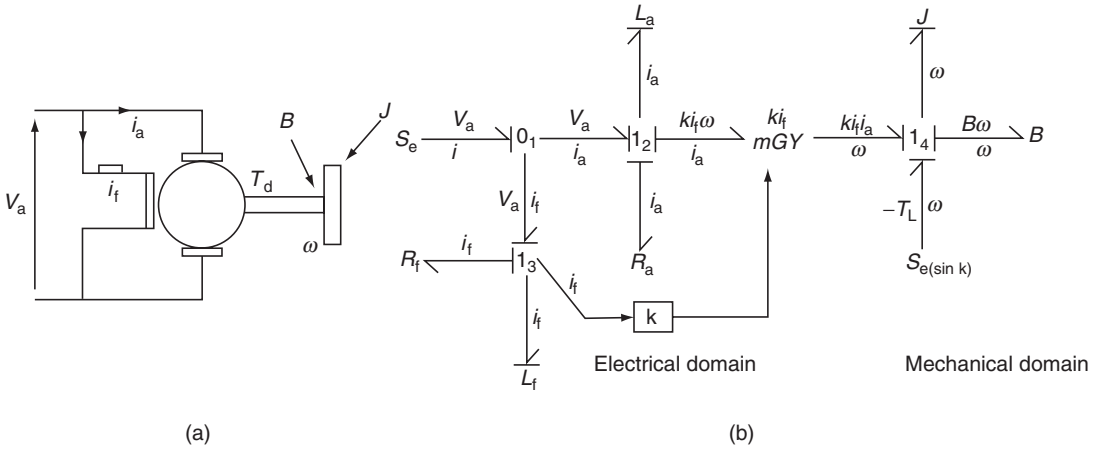


Figure 8.23 | (a) Armature excited DC motor; (b) bond graph with assigned causality.

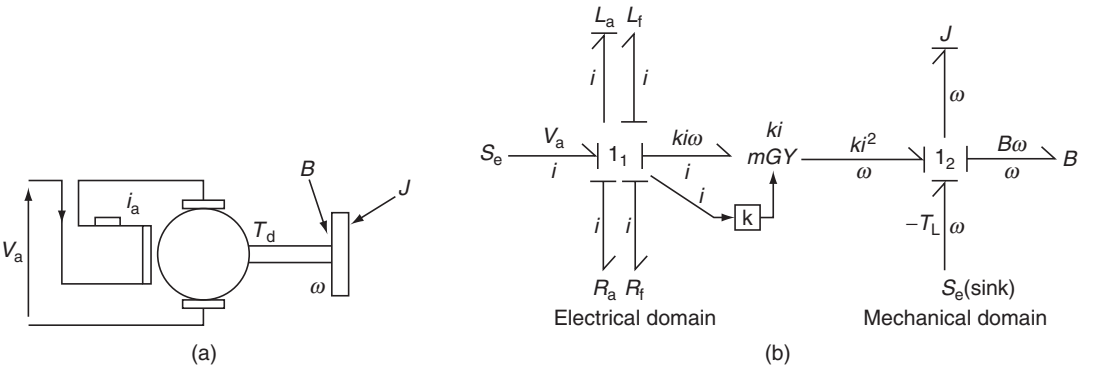


Figure 8.24 | (a) Series DC motor; (b) bond graph with assigned causality.

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{ki_f i_a}{J} - \frac{B\omega}{J} - \frac{T_L}{J} \tag{8.78b}$$

$$\frac{\mathrm{d}i_f}{\mathrm{d}t} = -\frac{R_f}{L_f} i_f + \frac{v}{L_a} \tag{8.78c}$$

DC Series Motor Figure 8.24(a) shows the schematic of a DC series motor wherein the field winding is in series with the armature. The bond graph of the DC series motor with the causality assigned is as shown in Figure 8.24(b). The series motor is also a non-linear system with state multiplication. The large-signal non-linear differential equations representing the DC series motor are given in Eqs. (8.79a) and (8.79b).

$$\frac{\mathrm{d}i}{\mathrm{d}t} = -\frac{R_a + R_f}{L_a + L_f} i - \frac{k}{L_a + L_f} i\omega + \frac{1}{L_a + L_f} v \tag{8.79a}$$

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{k}{J} i^2 - \frac{B}{J} \omega - \frac{T_L}{J} \tag{8.79b}$$

Modeling Switched Power Systems

Bond graphs have been used extensively in the modeling of physical systems. However, modeling switched power systems using bond graph pose some difficulties. This is because there is no switch element in bond graph. To overcome this problem, several authors have used macro models of power switches and several others have used modulated bond graph elements as a switch. The former method leads to very stiff systems and the latter method switches either the flow or effort variable losing control on the other variable. To overcome the problem of modeling switched systems, the concept of switched junctions is used. Here, the dominant junction variable is switched to emulate the practical switch. A few switched-mode power converters are modeled and simulated to demonstrate the efficacy of using the switched junctions.

Bond graphs have been used primarily for modeling dynamical systems where there is continuous energy or power flow through all components of the system. They do not inherently support modeling of switched power systems like switched-mode power converters and inverters. To overcome this problem, two approaches have generally been used: (a) one of the methods commonly used is to macro model the power

switches using inductances, capacitances and resistances; and (b) the other method is to use modulated transformers and gyrators with modulation index either being 1 or 0 depending on the state of the power switch. The first approach poses no problem in modeling power electronic systems wherein the components are operating in the linear region. However, for switched power electronic systems, the time constants of the macro models of the power switch will be in the order of microseconds and the simulation time will be in the order of seconds. Such a system wherein there is a large difference in time constants is called a stiff system. If such a system were to be simulated it will take days to finish. In the second approach wherein modulated transformers or gyrators may be used to represent power switches, the very nature of the effort–flow relationship may get altered. This approach controls either the effort or the flow of the power switch. As a consequence the model is not a correct representation of the physical system in most cases. To overcome the above-stated problems, the concept of switched junction is proposed which is a generalization of the conventional bond graph junction.

Switched Junctions

The 0-junction and the 1-junction are the multi-ports used in bond graph as discussed in a previous section. In this section the multi-ports are extended to a more generalized mathematical framework wherein multiple effort sources can decide a 0-junction effort and likewise multiple flow sources can decide a 1-junction flow.

The 0-junction is defined as

$$\sum_k f_k = 0$$

where f_k is the flow in the kth bond of the 0-junction. The bond effort is given as

$$e_k = e_{0j} \qquad \text{for all } k$$

where e_{0j} is the 0-junction effort. There is only one effort-decider bond at the junction which will determine the junction effort e_{0j}. Similarly, the 1-junction is defined as

$$\sum_k e_k = 0$$

where e_k is the effort in the kth bond of the 1-junction. The bond flow is given as

$$f_k = f_{1j} \qquad \text{for all } k$$

where f_{1j} is the 1-junction flow.

There is only one flow decider bond at the junction which will determine the junction flow, f_{1j}. The 0- and 1-junctions, as defined above are for continuous power flow through the system. To handle switched power flow through the system, these definitions are extended for the switched junctions. The *switched 0-junction or 0s-junction* is defined as follows:

$$\sum_k f_k = 0$$

where f_k is the flow in the kth bond of the 0-junction and

$$e_k = e_{0j} \qquad \text{for all } k$$

where e_{0j} is the 0s-junction effort. There can be many bonds connected to the 0s-junction with the causal bar at the junction end. This implies that the junction effort is decided by many bonds, however at mutually

exclusive instants of time. If e_i, f_i are the effort and flow, respectively, of the ith effort-decider bond of the switched 0s-junction and u_i is the information signal that selects the ith effort-decider bond as the active effort decider for the junction then

$$e_{0j} = u_i \cdot \begin{bmatrix} u_1 & u_2 & \cdots & u_n & u_{n+1} & \cdots \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \\ e_{i+1} \\ \vdots \end{bmatrix}$$

$$u_i \cdot u_n = \begin{cases} 0 & i \neq n \\ 1 & i = n \end{cases}$$

$$f_i = 0 \quad \text{if} \quad u_i \cdot u_n = 0$$

Note that this definition is a generalization of the 0-junction definition wherein there is only one effort decider bond at the 0-junction.

The *switched 1-junction or 1s-junction* is defined as follows:

$$\sum_k e_k = 0$$

where e_k is the effort in the kth bond of the 0-junction and

$$f_k = f_{1j} \qquad \forall\, k$$

where f_{1j} is the 1s-junction flow. There can be many bonds connected to the 1s-junction with the causal bar away from the junction end. This implies that the junction flow is decided by many bonds, however at mutually exclusive instants of time. If e_i, f_i are the effort and flow, respectively, of the ith flow decider bond of the switched 1s-junction and u_i is the information signal that selects the ith flow decider bond as the active flow decider for the junction then

$$f_{1j} = u_i \cdot \begin{bmatrix} u_1 & u_2 & \cdots & u_n & u_{n+1} & \cdots \end{bmatrix} \cdot \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_i \\ f_{i+1} \\ \vdots \end{bmatrix}$$

$$u_i \cdot u_n = \begin{cases} 0 & i \neq n \\ 1 & i = n \end{cases}$$

$$e_i = 0 \quad \text{if} \quad u_i \cdot u_n = 0$$

Note this definition is a generalization of the 1-junction definition wherein there is only one flow decider bond at the 1s-junction.

Application of Switched Junctions

The basic building block in a switched junction is a switch element. An SPDT switch is considered as the basic switch block. All other switch circuits can be built using many SPDT switches. The SPDT switch is shown in Figure 8.25. The SPDT switch consists of a pole P and two throw positions T1 and T2. The characteristic of this SPDT switch is that at the throw positions, the flow is being switched. When the pole P is switched between T1 and T2, the flow through T1 and T2 gets switched. Therefore, the throw positions of the switch will be represented by 1s-junction. The pole P of the switch will have switching in the effort and therefore, the pole position will be represented by a 0s-junction. When the pole P is at throw position T1, the flow through T2 is zero. The effort at the pole P is determined from the circuit connected to throw position T1. Similarly, when the pole P is at throw position T2, the flow through T1 is zero. The effort at the pole P is determined from the circuit connected to the throw position T2. The bond graph of the SPDT switch component is illustrated in Figure 8.26.

From Figure 8.26(b), one sees two 1s-junctions that are numbered 1 and 2 to represent the throw positions T1 and T2. There is one 0s-junction numbered 3 to represent the pole P. The throw positions T1 and T2 will take the flow of the pole (f) whenever the pole P is connected to them. The pole will take efforts of the throws (e_{T1} or e_{T2}) depending on whether the pole is connected to T1 or T2, respectively.

During the time when u is high or asserted, the pole effort e_p is taken as the effort e_{T1} which is from the circuit connected to throw T1. The throw flow f_{T1} is taken as f which is coming from the pole circuit. At this time the throw flow f_{T2} is zero which is coming from the zero flow source connected to the 1s-junction 2. During the time when $\bar{u}$ is asserted or high, the pole effort e_p is taken as the effort e_{T2} which is from the

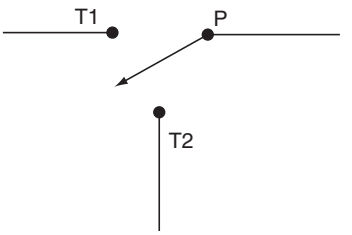


Figure 8.25 | Schematic of an SPDT switch.

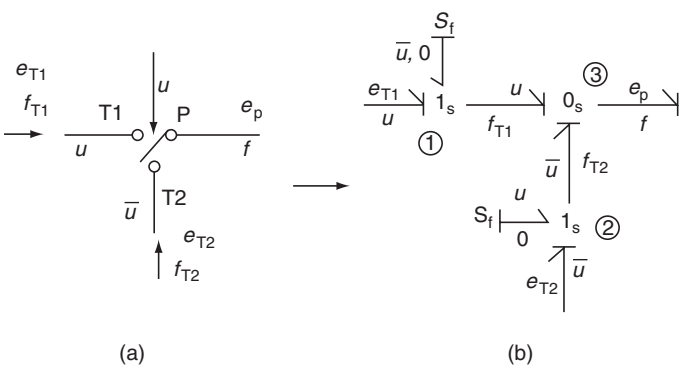


Figure 8.26 | Bond graph of an SPDT switch.

circuit connected to throw T2. The throw flow f_{T2} is taken as f which is coming from the pole circuit. Now f_{T1} is zero which is coming from the zero flow source connected to the 1s-junction 1.

Buck Converter

Consider the buck converter circuit shown in Figure 8.27(a) which is one of the primary switched-mode power converter. The bond graph of the buck converter circuit is shown in Figure 8.27(b). The state equations can now be obtained from the augmented bond graph based on the rules explained in the previous sections. However, one should note that the system is now a time-varying system as the equations are dependent on u which is a function of time.

Referring to Figure 8.27(b), the state equation can be obtained in a manner similar to that discussed in the previous section. There are two state variables: (a) the inductor current i_L and (b) the capacitor voltage v_C. Applying the "effort law" at the 1-junction to which the inductor is connected, one obtains

$$L\frac{\mathrm{d}i_L}{\mathrm{d}t} = u \cdot v_i - v_C \tag{8.80}$$

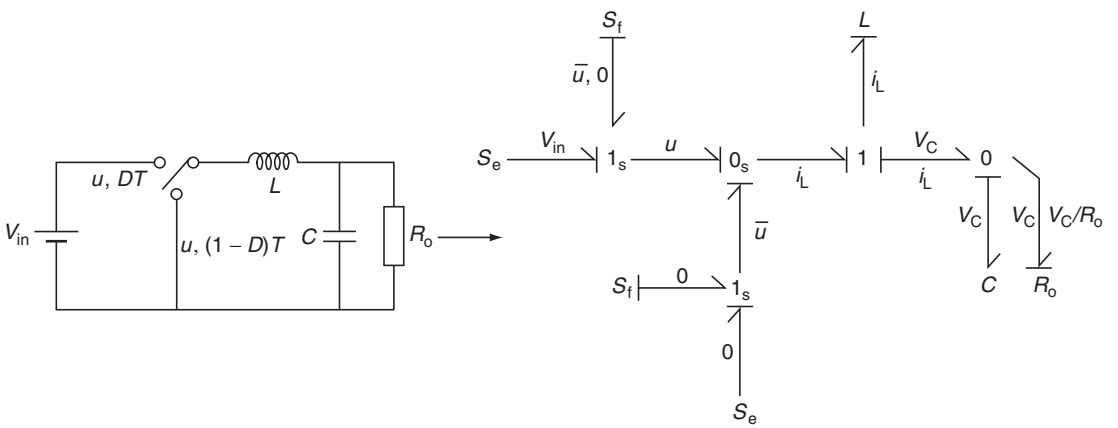
Applying the "flow law" at the 0-junction to which the capacitor is connected, we have

$$C\frac{\mathrm{d}v_C}{\mathrm{d}t} = i_L - \frac{v_C}{R_o} \tag{8.81}$$

Equations (8.80) and (8.81) can be re-arranged in the standard form to obtain

$$\begin{bmatrix} \dfrac{\mathrm{d}i_L}{\mathrm{d}t} \\ \dfrac{\mathrm{d}v_C}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} 0 & \dfrac{-1}{L} \\ \dfrac{1}{C} & \dfrac{-1}{R_o C} \end{bmatrix} \cdot \begin{bmatrix} i_L \\ v_C \end{bmatrix} + \begin{bmatrix} \dfrac{u}{L} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} v_i \end{bmatrix} \tag{8.82}$$

Equation (8.82) is the large-signal model of the buck converter. Observe that the *input* matrix contains the Boolean variable u. During interval dT_s, when the switch pole is connected to the input, $u = 1$. During



(a) (b)

Figure 8.27 | (a) Buck converter circuit; (b) augmented bond graph of buck converter.

$(1 - d)T_s$ interval, the switch pole is connected to ground and $u = 0$. From Eq. (8.82), the state equations for the two operative positions of the switch are obtained by substituting $u = 1$ and $u = 0$ for the two intervals. This will result in the same large-signal model as obtained with the circuit averaging method. Once the large-signal model is obtained, the averaged large-signal, steady-state and the small-signal models are obtained in the same manner as that discussed in the section on circuit averaging method.

Boost Converter

Figure 8.28 shows the circuit topology of a boost converter circuit which is operating in continuous conduction mode (CCM). The operation and steady-state analysis of the boost converter circuit is discussed

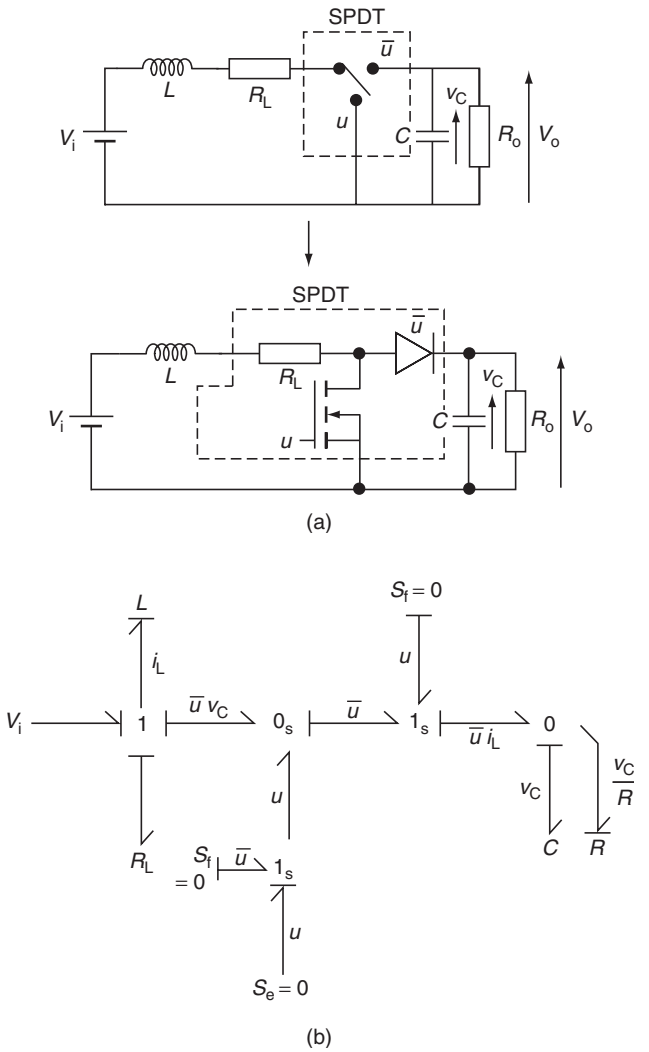


Figure 8.28 | (a) Boost converter; (b) bond graph of the boost converter.

in Chapter 5. The SPDT switch is realized with a MOSFET and a diode as shown in the figure. The bond graph of the boost converter is shown in Figure 8.28. The SPDT switch model of Figure 8.26 is used to represent the MOSFET-diode SPDT switch combination.

The 0s-junction has two effort-decider bonds which show that there are only two physically feasible operating modes for this configuration. Applying the "effort law" for the 1-junction to which the inductor is connected, one obtains

$$L\frac{\mathrm{d}i_{\mathrm{L}}}{\mathrm{d}t} = v_{\mathrm{i}} - i_{\mathrm{L}}R_{\mathrm{L}} - \overline{u}\cdot v_{\mathrm{C}} \tag{8.83}$$

Applying the "flow law" for the 0-junction to which the capacitor is connected, one obtains

$$C\frac{\mathrm{d}v_{\mathrm{C}}}{\mathrm{d}t} = \overline{u}\cdot i_{\mathrm{L}} - \frac{v_{\mathrm{C}}}{R_{\mathrm{o}}} \tag{8.84}$$

From Eqs. (8.83) and (8.84), the large-signal model can be obtained as

$$\begin{bmatrix} \dfrac{\mathrm{d}i_{\mathrm{L}}}{\mathrm{d}t} \\ \dfrac{\mathrm{d}v_{\mathrm{C}}}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} \dfrac{-R_{\mathrm{L}}}{L} & \dfrac{-\overline{u}}{L} \\ \dfrac{\overline{u}}{C} & \dfrac{-1}{R_{\mathrm{o}}C} \end{bmatrix} \cdot \begin{bmatrix} i_{\mathrm{L}} \\ v_{\mathrm{C}} \end{bmatrix} + \begin{bmatrix} \dfrac{1}{L} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} v_{\mathrm{i}} \end{bmatrix} \tag{8.85}$$

The Boolean variable $\overline{u} = 0$ during the interval dT_{s} and $\overline{u} = 1$ during the period $(1 - d)T_{\mathrm{s}}$. Substituting for $\overline{u}$ accordingly, the large-signal models for the two operative modes of the SPDT switch are obtained. Once the large-signal model is obtained, the averaged large-signal, steady-state and the small-signal models are obtained in the same manner as that discussed in the section on circuit averaging method.

8.9 Space-Vector Modeling

The electrical machines are electromechanical devices wherein there are three energy domains, namely electrical, magnetic and mechanical domains. The electrical power sources for these electrical machines are, in general, obtained from three-phase grid systems. A rotating field is produced in almost all of these electromechanical devices to produce the motive torque. Systems that are energized by three- or poly-phase energy sources can be easily modeled by using *space vectors*. The space-vector modeling approach is a generic method that can be used for two-, three- or poly-phase systems. However, this section will give a flavor for space-vector modeling by considering a three-phase induction motor as a continuous illustrative example in this section.

The operation of the three-phase induction motor is based on Faraday's law and the Lorentz force on a conductor. According to Faraday's law, if the flux linking a conductive loop varies as a function of time, then an emf which is proportional to the rate of change of flux is induced in the loop, thereby forcing a current to flow in the conductive loop. In the case of the induction motor, the induced currents in the rotor conductors interact with the air-gap flux to produce the torque. As the rotor tries to catch up with the magnetic flux in accordance with Lenz's law, the rate at which the conductors are cut by the magnetic flux is reduced. As a consequence, the induced rotor currents decrease and the Lorentz force on the rotor conductors reduces. Evidently, if the rotor conductors were to catch up with the magnetic flux, there would be no relative motion between the conductors and the flux, and this would result in zero rotor currents and zero torque.

Therefore, for the induction motor to operate, there must exist some relative speed between the flux and the conductors which is called the slip speed ω_{sl}. The slip speed is given by

$$\omega_{sl} = \omega_s - \omega_m \tag{8.86}$$

where ω_s is the synchronous speed and ω_m is the electrical rotor speed.

Rotating Magnetic Field

It is essential that a rotating magnetic field be produced for the rotation of the induction motor. Consider the schematic of a simple two-pole induction machine as shown in Figure 8.29. Three identical windings AN, BN, CN each composed of two coils (Aa–aN, etc.) in series are arranged to be spatially spaced at 120° to each other. With respect to the terminals A, B and C, the windings constitute a balanced three-phase system. The spatial representation of the three-phase system is shown in Figure 8.29(b). The three-phase system is considered to have three spatial axes "*sa*", "*sb*" and "*sc*" which are displaced at 120° with respect to each other.

Note that i_{sa}, i_{sb} and i_{sc} produce the three-phase stator flux linkages ψ_{sa}, ψ_{sb} and ψ_{sc}. When three-phase sinusoidal currents are fed to a balanced three-phase induction motor, a rotating magnetic field is generated. It should be noted that the stator windings should be distributed in such a manner which results in three-phase sinusoidal flux linkages on application of the three-phase sinusoidal stator currents. To handle three-phase quantities by means of a single equivalent vector quantity, the concept of space vectors is introduced in the next sub-section.

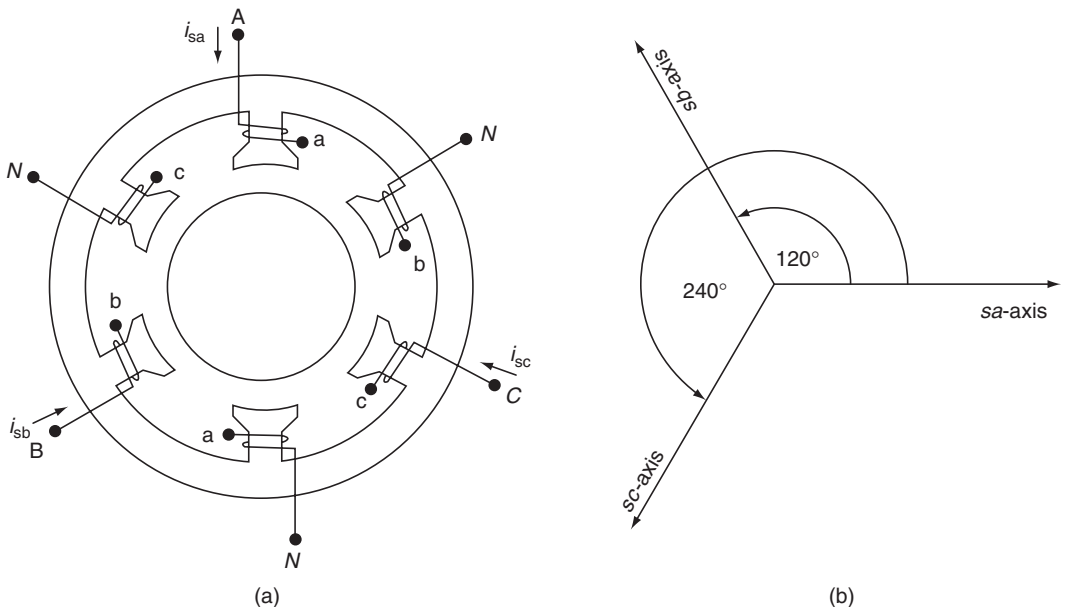


Figure 8.29 (a) Stator structure showing the three-phase windings; (b) spatial representation of *a*, *b* and *c* axes.

Space Vectors

The induction motor is a three-phase machine. Therefore, all its variables like the stator currents, rotor currents, etc. are three-phase quantities each having three components. Consider for the moment the stator current components i_{sa}, i_{sb} and i_{sc}. Referring to Figure 8.30, one can note that the i_{sa} component of the stator current is along an axis in space which is represented by the *sa*-axis. Similarly, the i_{sb} and i_{sc} components of the stator current are represented by the *sb*-axis and *sc*-axis in space, respectively. As these current components are represented as vectors in the spatial co-ordinates, they are called current space vectors. Considering the *sa*-axis as the spatial reference, it is evident from Figure 8.30 that the *sb*-axis and the *sc*-axis are positioned spatially at $2\pi/3$ and $4\pi/3$ radians, respectively, measured anti-clockwise from the *sa*-axis. The *b*-phase and *c*-phase stator current space vectors can be represented with respect to the *sa*-axis as

$$i_{sb} = i_{sb}e^{j(2\pi/3)};\ \ i_{sc} = i_{sc}e^{j(4\pi/3)} \tag{8.87}$$

The three stator current components can be represented by a single equivalent vector, called the resultant stator current space vector i_s, which is given by

$$i_s = i_{sa} + i_{sb}e^{j(2\pi/3)} + i_{sc}e^{j(4\pi/3)} \tag{8.88}$$

Similarly, any three-phase variable in the induction motor can be represented as a resultant space vector. If χ represents a general three-phase variable of the induction motor, then the resultant space vector of the variable χ is given by

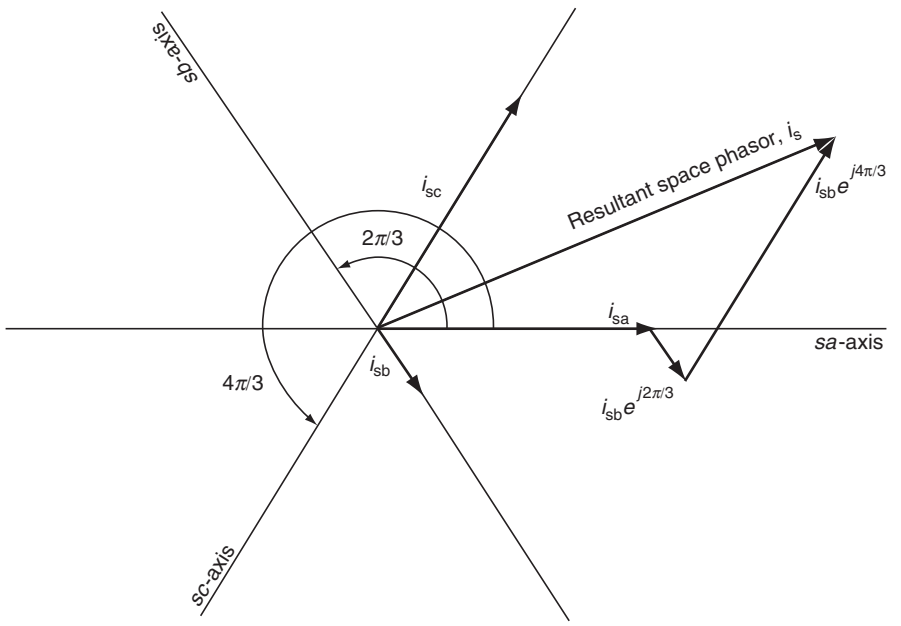


Figure 8.30 | Space-vector representation of stator currents of the induction motor.

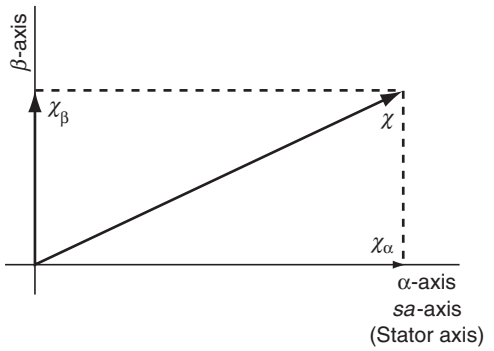


Figure 8.31 | Representation of a general space phasor along the orthogonal axes.

$$\chi = \chi_a + \chi_b \cdot e^{j(2\pi/3)} + \chi_c \cdot e^{j(4\pi/3)} \tag{8.89}$$

where the variable χ can represent the currents i_s and i_r; voltages, V_s,V_r; fluxes, ϕ_s, ϕ_r, ϕ_m and flux linkages ψ_s, ψ_r, ψ_m. It should be noted clearly that χ_a, χ_b and χ_c denote phase quantities and not line quantities.

Representation of Space Vectors in Orthogonal Co-ordinates

The resultant space vector χ can be represented with respect to two orthogonal axes, α and β. The α-axis is aligned along the *sa*-axis as shown in Figure 8.31. The projections of the space vector χ along the α-axis and β-axis are χ_α and χ_β, respectively, as shown in Figure 8.31. These orthogonal space-vectors components are simple transformation as given by

$$\begin{bmatrix} \chi_\alpha \\ \chi_\beta \end{bmatrix} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix} \cdot \begin{bmatrix} \chi_a \\ \chi_b \\ \chi_c \end{bmatrix} \tag{8.90}$$

Space-Vector Transformations

Referring to Figure 8.32, the α-axis and the β-axis are fixed in space. If the space vector χ rotates, then the projections χ_α and χ_β of χ on the two orthogonal axes will vary even if the magnitude of χ is a constant. This is clearly shown in Figure 8.32(a), where the space vector χ shown at two instants of time t_1 and t_2 results in different projections along the α–β axis, which is stationary. However, if the α–β axis also rotates along with the space vector at the same speed, then the projections χ_α and χ_β on the rotating α–β axis will remain constant if the magnitude of χ is a constant as shown in Figure 8.32(b).

As an illustration, the three-phase stator currents of an induction motor, under steady-state operating conditions, are sinusoidal quantities. The three-phase stator currents (i_{sa}, i_{sb} and i_{sc}) are given by the expression

$$i_{sa} = I \sin\omega t;\ i_{sb} = I \sin(\omega t - 2\pi/3);\ i_{sc} = I \sin(\omega t - 4\pi/3)$$

The stator current space vector i_s, as given by Eq. (8.88), reduces to $i_s = (3I/2)\ e^{j(\omega t - \pi/2)}$. The resolved components $i_{s\alpha}$ and $i_{s\beta}$ of i_s in the α–β reference frame are given by Eq. (8.90). Thus, $i_{s\alpha} = (3I/2)\cos(\omega t - \pi/2)$ and $i_{s\beta} = (3I/2)\sin(\omega t - \pi/2)$. It is evident that the magnitude of the space vector is $3I/2$, which is a constant and the space vector is rotating at speed ω. If the α–β axis is fixed in space, then the projections of

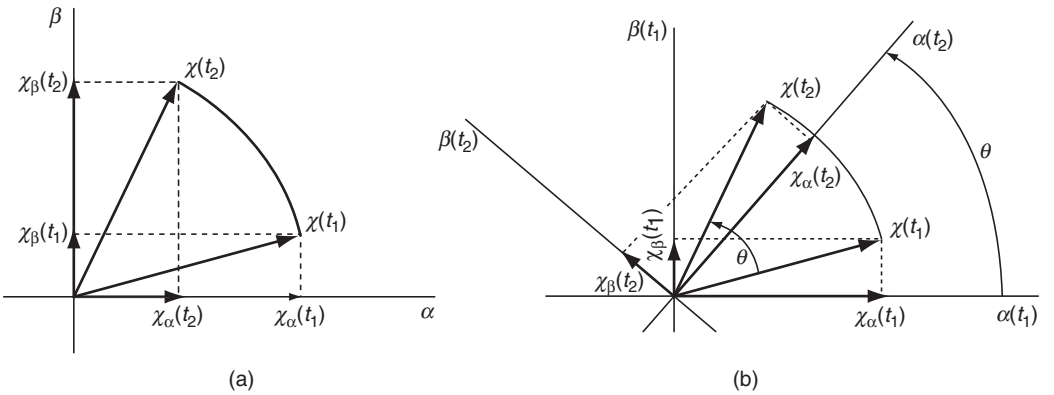


Figure 8.32 | (a) Stationary α–β reference frame; (b) α–β reference frame rotating synchronously with χ.

the stator current space vector $i_{s\alpha}$ and $i_{s\beta}$ along the stationary α–β axis will vary sinusoidally. However, if the α–β axis also rotates at the same speed ω as that of the space vector i_s, then the resolved components of the stator current space vector will be DC in nature. Thus, even if the motor sees sinusoidally varying currents, the resolved components of the currents along the rotating orthogonal axes are unchanging with time, that is, are DC quantities.

This concept is exploited in the vector control of induction motors. (However, one should note that non-sinusoidal quantities in the stationary reference frame will result in a varying unipolar quantity in a reference frame rotating synchronously with the space vector – this is the case during transient conditions in the induction motor.)

Figure 8.33 shows the graphical representation of a space vector r. This has a magnitude r and is at an angle θ from the x-axis. This is represented in the polar co-ordinates as follows:

$$r_{x\text{-axis}} = r \cdot e^{j\theta} \tag{8.91}$$

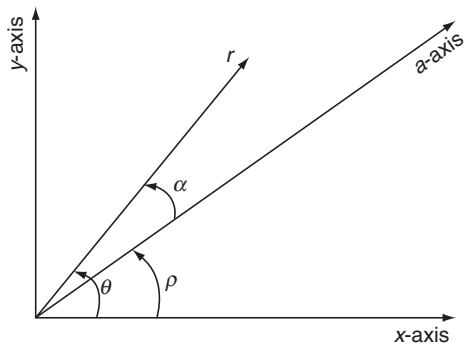


Figure 8.33 | Space-vector representation in polar co-ordinates.

The space vector r is to be represented with respect to an arbitrary reference axis (a-axis) which is at an angle ρ with respect to the x-axis as shown in Figure 8.33. (It can be noted that the arbitrary axis may or may not rotate synchronously with the space vector r.) The space vector r can now be represented as

$$r_{\text{a-axis}} = r \cdot e^{ja} = r \cdot e^{j(\theta-\rho)} = r_{\text{x-axis}} \cdot e^{-j\rho} \tag{8.92}$$

From Eq. (8.92), it is evident that to represent a space vector r with respect to an arbitrary reference axis (a-axis), the x-axis representation is multiplied by a factor $e^{-j\rho}$ which is called the *transformation factor*. Generalizing, one may represent any space vector with respect to a new reference axis by multiplying the representation of the space vector with respect to the old reference axis by the transformation factor $e^{-j\rho}$, where ρ is the relative shift of the new axis with respect to the old axis.

α and ρ are to be taken as positive when measured in the anti-clockwise direction and negative when measured in the clockwise direction. The arbitrary axis (a-axis) is called the *d-axis* or direct axis and the axis perpendicular to the arbitrary axis is called the *q-axis* or quadrature axis.

In Figure 8.34, an induction motor space vector χ is shown resolved along two co-ordinate systems, one with respect to the α–β axes and the other with respect to the d–q axes. The d–q co-ordinate system is phase shifted with respect to the α–β co-ordinate system by an angle ρ. Applying the transformation as indicated in Eq. (8.92) and splitting into real and imaginary parts, one obtains

$$\begin{bmatrix} \chi_{\text{d}} \\ \chi_{\text{q}} \end{bmatrix} = \begin{bmatrix} \cos\rho & \sin\rho \\ -\sin\rho & \cos\rho \end{bmatrix} \cdot \begin{bmatrix} \chi_{\alpha} \\ \chi_{\beta} \end{bmatrix} \tag{8.93}$$

Equation (8.93) gives the transformation for transforming the two-phase variables in the α–β co-ordinates to the two-phase variables in the d–q co-ordinates. If the d–q co-ordinate system is rotating synchronously with the space vector χ, then the χ_{d} and χ_{q} components will be constants if the magnitude of χ is a constant. It can be noted that in the context of the transformations carried out in this section, the α–β co-ordinates form the *stator reference frame* and the d–q co-ordinates form the *arbitrary reference frame*, as explained earlier. In the next sub-section, modeling of the induction motor using the concepts of the space vectors will be discussed.

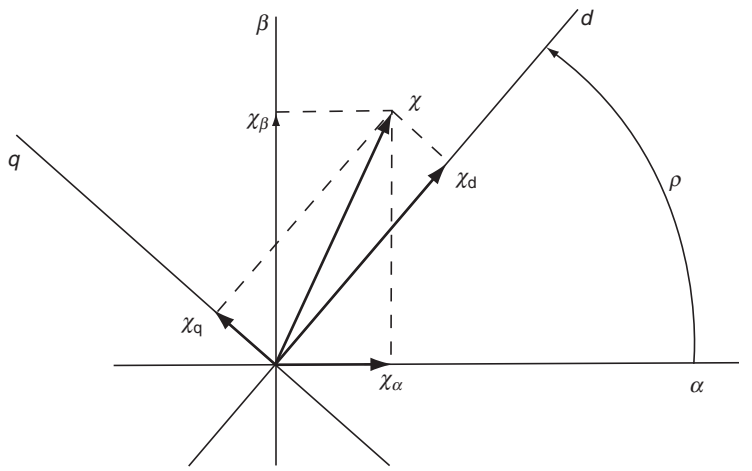


Figure 8.34 | Space-vector representation in stator (α–β) and arbitrary (d–q) reference frames.

Modeling of Induction Motor

In developing the dynamic model of the induction motor, the following assumptions are made:

1. The motor has symmetrical three-phase three-wire windings.
2. The mmf wave is sinusoidally distributed in space.
3. The stator and rotor iron have infinite permeability.
4. Skin effect and core losses are negligible.
5. The motor is operating in the linear region of the *B–H* characteristics of the stator and the rotor materials.

The voltage which is supplied to the stator of the induction motor should

1. sustain the drop across the stator resistance;
2. develop the emf required for setting up the stator flux linkage.

As a result

$$v_s = i_s \cdot R_s + \frac{d\psi_s}{dt} \tag{8.94}$$

where v_s, i_s and ψ_s are space vectors as defined by Eq. (8.89). Similarly, the rotor voltage v_r, expressed with respect to the rotor reference frame, is given by

$$v_r = i_r \cdot R_r + \frac{d\psi_r}{dt} \tag{8.95}$$

where v_r, i_r and ψ_r are space vectors as defined by Eq. (8.89). The stator and the rotor flux linkages can be expressed in terms of the stator and the rotor current space vectors as follows:

$$\psi_s = L_{ss} \cdot i_s + M \cdot i_r \cdot e^{j\varepsilon} \tag{8.96}$$

$$\psi_r = M \cdot i_s + L_{rr} \cdot i_r \cdot e^{j\varepsilon} \tag{8.97}$$

where

$$L_{ss} = L_{\sigma s} + M \tag{8.98}$$

$$L_{rr} = L_{\sigma r} + M \tag{8.99}$$

Also $L_{\sigma s}$, $L_{\sigma r}$ are the per phase leakage inductances of the stator and rotor, respectively; M is the equivalent three-phase magnetizing inductance. In Eqs. (8.96) and (8.97) the space vectors ψ_s, ψ_r and i_s are represented in the stator reference frame. But the rotor current space vector i_r is represented with respect to the rotor axis which is at an angle of ε relative to the stator axis (see Figure 8.35). Therefore the rotor current space vector i_r in Eqs. (8.96) and (8.97) is multiplied by the transformation factor $e^{j\varepsilon}$ (as explained in the previous sub-section) so that all the variables in Eq. (8.96) and (8.97) are represented in the stator reference frame.

Substituting ψ_s and ψ_r from Eqs. (8.96) and (8.97) into Eqs. (8.94) and (8.95), respectively, the induction motor voltage equations in the *stator reference frame* are given by

$$v_s = R_s \cdot i_s + L_{ss} \cdot \frac{di_s}{dt} + M \cdot \frac{d(i_r \cdot e^{is})}{dt} \tag{8.100}$$

$$v_r \cdot e^{j\varepsilon} = R_r \cdot i_r \cdot e^{j\varepsilon} + L_{rr} \cdot e^{j\varepsilon} \cdot \frac{di_r}{dt} + M \cdot e^{j\varepsilon} \cdot \frac{d(i_s \cdot e^{-js})}{dt} \tag{8.101}$$

The dynamics of the rotor shaft and the load is governed by

$$\frac{J}{p}\frac{\mathrm{d}\omega_{\mathrm{m}}}{\mathrm{d}t} = T_{\mathrm{d}} - T_{\mathrm{L}} - \frac{B}{p}\omega_{\mathrm{m}} \tag{8.102}$$

where T_{d} is the electromagnetic drive torque; T_{L} the load torque; B the friction coefficient; J the inertia seen on the rotor shaft; p the number of pole pairs; ω_{m} the electrical rotor speed which is "p" times the shaft speed. Equations (8.100)–(8.102) represent the model of the induction motor. In the following sub-section, the model is represented in the arbitrary reference frame to arrive at the *d–q* axes model of the induction motor.

Representation of the Voltages in the Arbitrary Reference Frame

Consider an arbitrary reference frame as shown in Figure 8.35 that is rotating at speed ω_{a} (which is not necessarily equal to the synchronous speed ω_{s}). Let ρ be the angle between the stator reference axis α and *d*-axis at any given instant of time. Then $\mathrm{d}\rho/\mathrm{d}t = \omega_{\mathrm{a}}$. The voltage equations [Eqs. (8.100) and (8.101)] in the stator reference frame can be transformed to the arbitrary reference frame by using the co-ordinate transformation factor $e^{-j\rho}$. Ultimately, the intention is to ensure that the arbitrary reference frame is rotating at the synchronous speed, that is, $\omega_{\mathrm{a}} = \omega_{\mathrm{s}}$, so that all the variables appear as DC quantities under steady-state operating conditions. Therefore,

$$v_{\mathrm{s}} \cdot e^{-j\rho} = R_{\mathrm{s}} \cdot i_{\mathrm{s}} \cdot e^{-j\rho} + L_{\mathrm{ss}} \cdot e^{-j\rho} \cdot \frac{\mathrm{d}i_{\mathrm{s}}}{\mathrm{d}t} + M \cdot e^{-j\rho} \cdot \frac{\mathrm{d}(i_{\mathrm{r}} \cdot e^{js})}{\mathrm{d}t} \tag{8.103}$$

$$v_{\mathrm{r}} \cdot e^{j(\varepsilon - p)} = R_{\mathrm{r}} \cdot i_{\mathrm{r}} \cdot e^{j(\varepsilon-\rho)} + L_{\mathrm{rr}} \cdot e^{j(\varepsilon-\rho)} \cdot \frac{\mathrm{d}i_{\mathrm{r}}}{\mathrm{d}t} + M \cdot e^{j(\varepsilon-\rho)} \cdot \frac{\mathrm{d}(i_{\mathrm{s}} \cdot e^{-js})}{\mathrm{d}t} \tag{8.104}$$

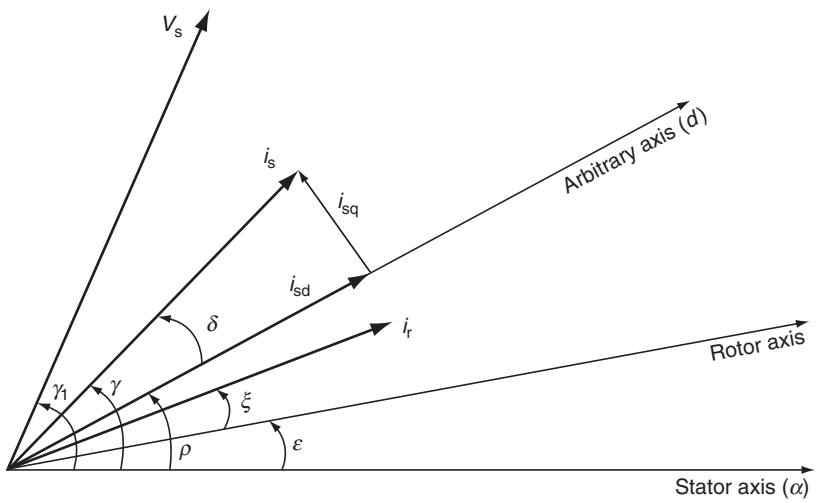


Figure 8.35 | Space-vector diagram for the induction motor.

Referring to Figure 8.35, the stator current and voltage space vectors i_s and v_s represented with respect to the stator reference frame are given by

$$i_s = i_s \cdot e^{j\gamma} \tag{8.105}$$

$$v_s = v_s \cdot e^{j\gamma_1} \tag{8.106}$$

Similarly, the rotor current and the voltage space vectors i_r and v_r represented with respect to the rotor axis are given by

$$i_r = i_r \cdot e^{j\xi} \tag{8.107}$$

$$v_r = v_r \cdot e^{j\xi_1} \tag{8.108}$$

Substitution of Eqs. (8.105)–(8.108) in Eqs. (8.103) and (8.104), results in

$$v_s \cdot e^{j(\gamma_1-\rho)} = R_s \cdot i_s \cdot e^{j(\gamma-\rho)} + L_{ss} \cdot e^{-j\rho} \cdot \frac{\mathrm{d}(i_s \cdot e^{j\gamma})}{\mathrm{d}t} + M \cdot e^{-j\rho} \cdot \frac{\mathrm{d}(i_r e^{j(\xi+s)})}{\mathrm{d}t} \tag{8.109}$$

$$v_r \cdot e^{j(\xi_1+\varepsilon-\rho)} = R_r \cdot i_r \cdot e^{j(\xi+\varepsilon-\rho)} + L_{rr} \cdot e^{j(\varepsilon-\rho)} \cdot \frac{\mathrm{d}(i_r \cdot e^{j\xi})}{\mathrm{d}t} + M \cdot e^{j(\varepsilon-\rho)} \cdot \frac{\mathrm{d}(i_s \cdot e^{j(\gamma-s)})}{\mathrm{d}t} \tag{8.110}$$

For a squirrel cage induction motor, $v_r = 0$ as the rotor conductors are short circuited. Referring to Figure 8.34, one can define the following relationships:

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = \omega_m \text{ (the mechanical speed of the induction motor)} \tag{8.111}$$

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \omega_\xi = \omega_{sl} \text{ (the slip speed of the induction motor)} \tag{8.112}$$

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \omega_s \text{ (synchronous speed)} \tag{8.113}$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \omega_a \text{ (speed of the arbitrary reference frame)} \tag{8.114}$$

The d–q Axes Model

The voltage space-vector equations [Eqs. (8.109) and (8.110)] are split into two orthogonal components to obtain the *d–q* axes model of the induction motor. As any space vector can be represented as projections of the space-vector magnitude along any orthogonal co-ordinate system, one can define the following relationships with respect to the arbitrary reference frame:

Direct axis stator voltage, $V_{sd} = v_s \cos(\gamma_1 - \rho)$ (8.115a)

Quadrature axis stator voltage, $V_{sq} = v_s \sin(\gamma_1 - \rho)$ (8.115b)

Direct axis stator current, $i_{sd} = i_s \cos(\gamma - \rho)$ (8.115c)

Quadrature axis stator current, $i_{sq} = i_s \sin(\gamma - \rho)$ (8.115d)

$$\text{Direct axis rotor current, } i_{rd} = i_r \cos(\xi + \varepsilon - \rho) \tag{8.115e}$$

$$\text{Quadrature axis rotor current, } i_{rq} = i_r \sin(\xi + \varepsilon - \rho) \tag{8.115f}$$

Using the definitions given in Eqs. (8.111)–(8.115), the stator voltage space-vector equations [Eqs. (8.109) and (8.110)] can be represented as stator and rotor voltage equations along the direct and the quadrature axes of an arbitrary reference frame. Thus

$$V_{sd} = R_s \cdot i_{sd} + L_{ss} \cdot \frac{di_{sd}}{dt} - L_{ss} \cdot \omega_a \cdot i_{sq} + M \cdot \frac{di_{rd}}{dt} - M \cdot \omega_a \cdot i_{rq} \tag{8.116}$$

$$V_{sq} = R_s \cdot i_{sq} + L_{ss} \cdot \frac{di_{sq}}{dt} + L_{ss} \cdot \omega_a \cdot i_{sd} + M \cdot \frac{di_{rq}}{dt} + M \cdot \omega_a \cdot i_{rd} \tag{8.117}$$

$$0 = R_r \cdot i_{rd} + L_{rr} \cdot \frac{di_{rd}}{dt} - L_{rr} \cdot (\omega_a - \omega_m) \cdot i_{rq} + M \cdot \frac{di_{sd}}{dt} - M \cdot (\omega_a - \omega_m) \cdot i_{sq} \tag{8.118}$$

$$0 = R_r \cdot i_{rq} + L_{rr} \cdot \frac{di_{rq}}{dt} + L_{rr} \cdot (\omega_a - \omega_m) \cdot i_{rd} + M \cdot \frac{di_{sq}}{dt} + M \cdot (\omega_a - \omega_m) \cdot i_{sd} \tag{8.119}$$

Equations (8.116)–(8.119) can be represented in matrix form as follows:

$$\begin{bmatrix} V_{sd} \\ V_{sq} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \cdot \begin{bmatrix} i_{sd} \\ i_{sq} \\ i_{rd} \\ i_{rq} \end{bmatrix} \tag{8.120}$$

where

$$a_{11} = a_{22} = R_s + L_{ss} \cdot \frac{d}{dt}; \quad a_{33} = a_{44} = R_r + L_{rr} \cdot \frac{d}{dt}$$

$$a_{12} = -a_{21} = -L_{ss} \cdot \omega_a; \quad a_{13} = a_{24} = a_{31} = a_{42} = M \cdot \frac{d}{dt}$$

$$a_{14} = -a_{23} = -M \cdot \omega_a; \quad a_{32} = -a_{41} = -M \cdot (\omega_a - \omega_m)$$

$$a_{34} = -a_{43} = -L_{rr} \cdot (\omega_a - \omega_m)$$

Equations (8.116)–(8.119) are called the *d–q* axes voltage equations of the induction motor which results in the *d–q* axis equivalent circuit (in the arbitrary reference frame) as shown in Figure 8.36. The flux linkages shown in the equivalent circuit of Figure 8.36 are given by

$$\psi_{sd} = L_{ss} i_{sd} + M i_{rd} \tag{8.121}$$

$$\psi_{sq} = L_{ss} i_{sq} + M i_{rq} \tag{8.122}$$

$$\psi_{rd} = M i_{sd} + L_{rr} i_{rd} \tag{8.123}$$

$$\psi_{rq} = M i_{sq} + L_{rr} i_{rq} \tag{8.124}$$

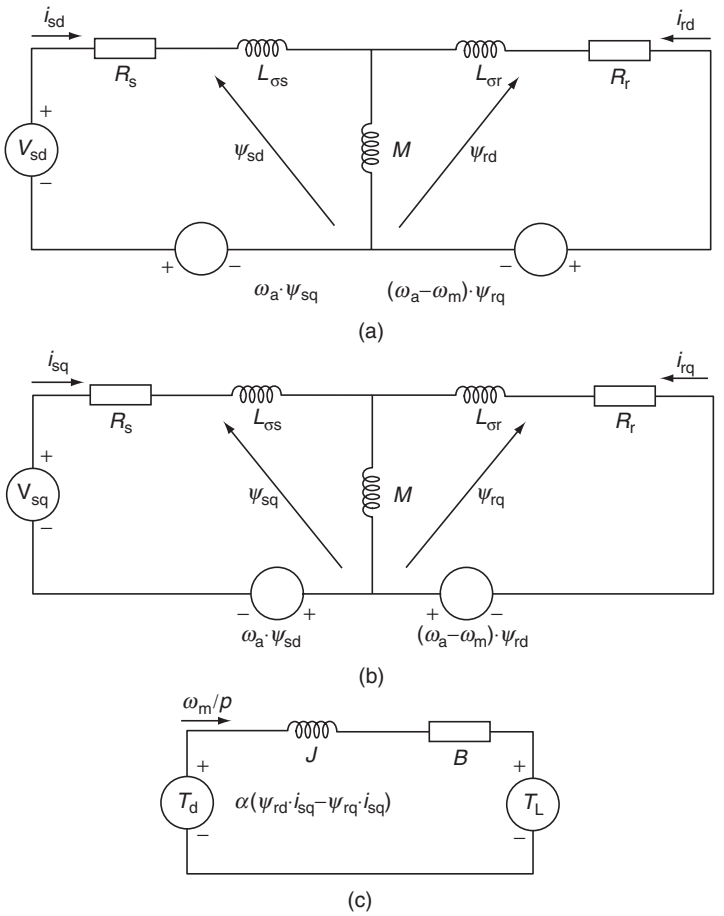


Figure 8.36 The *d–q* axis equivalent circuit of the induction motor: (a) Direct axis; (b) quadrature axis; (c) torque model w.r.t. rotor fluxes.

When $\omega_a = \omega_s$, this equivalent circuit is with respect to the synchronous reference frame, and when $\omega_a = 0$, then this equivalent circuit is with respect to the stator reference frame. Equations (8.120) and (8.102) together form the model of the induction motor. This is generally referred to as the *d–q* model of the induction motor where a three-phase induction motor is represented as an equivalent two-phase (*d*-phase and *q*-phase) motor. In the next sub-section, the state space approach to modeling is presented, as it is preferred for controller and flux estimator synthesis.

State Space Representation of the d–q Model of the Induction Motor

The model of the induction motor as represented in Eq. (8.120) does not give the induction motor model in the form of the standard state equation form. Therefore the various dynamics involved with respect to the stator currents and the flux linkages are not very clear at the outset. For the design of the controller, it is essential to clearly understand the nature of the time constants involved with respect to the stator currents

and the flux linkages. Therefore it is necessary to represent the *d–q* model of the induction motor given by Eqs. (8.120) and (8.102) in the standard *state space form*. In the state space representation, the various time constants and the interaction between the various state variables can be clearly visualized. One can note that in the state space description of the model, the *A*, *B*, *C* and *D* matrices will be free of time derivatives for a time-invariant system. The state space representation of the *d–q* model of the induction motor is essential for synthesis of flux estimators and controllers.

From the model of the induction motor discussed in the previous section, it is evident that the induction motor system can be split into two dynamical systems. One of the systems consists of the torque dynamics which involves Eq. (8.120) and the other system consists of the speed dynamics which involves Eq. (8.102). These two dynamical sub-systems are shown in Figure 8.37. The state space model for the speed dynamics is directly obtainable from Eq. (8.102) which leads to

$$\frac{\mathrm{d}(\omega_\mathrm{m} / p)}{\mathrm{d}t} = \left[-\frac{B}{J}\right] \cdot \frac{\omega_\mathrm{m}}{p} + \left[\frac{1}{J}\right] \cdot (T_\mathrm{d} - T_\mathrm{L}) \tag{8.125}$$

where (ω_m/p) is the state variable for the speed dynamics indicated in Eq. (8.125). To obtain the state space model for the torque dynamics, Eq. (8.120) is used in conjunction with Eqs. (8.121)–(8.124). The resulting equations can be reduced to the conventional state space form given by

$$\dot{x} = A \cdot x + B \cdot u \tag{8.126}$$

$$y = C \cdot x \tag{8.127}$$

where x is the state vector (stator currents and flux linkages); u the input vector (stator voltages) and y the output vector (stator currents). The exact state variables depend on the flux linkage of interest.

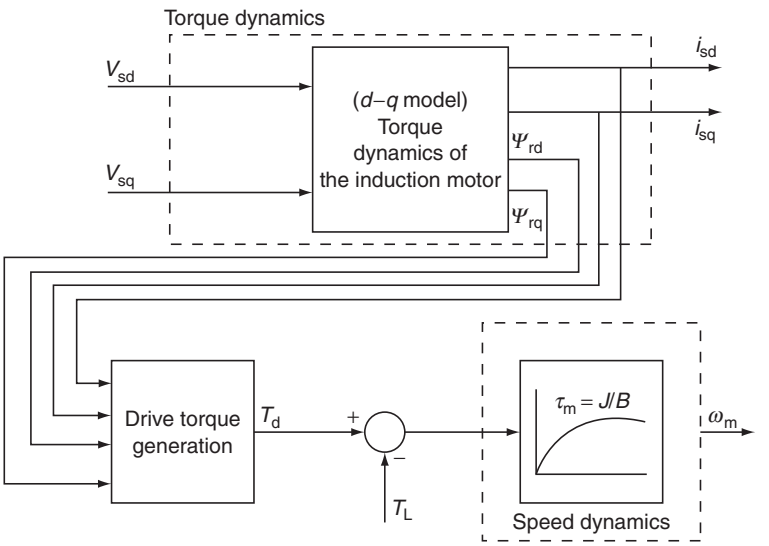


Figure 8.37 | Induction motor system split into torque and speed dynamical sub-systems.

While considering the state space models with respect to the arbitrary reference frame the following models can be obtained depending on the choice of the state variables:

1. rotor flux state model;
2. stator flux state model;
3. air-gap flux state model.

In addition, a concept of *generalized flux state model* will be introduced which can represent any of the above three flux state models.

Rotor Flux State Model

In this model, the state vector consists of the stator currents and the rotor flux linkages as state variables. The different vectors used in this model are defined as follows:

State vector $x = [i_{sd}, i_{sq}, \psi_{rd}, \psi_{rq}]^T$

Input vector $u = [V_{sd}, V_{sq}]^T$

Output vector $y = [i_{sd}, i_{sq}]^T$

The system matrix A is derived from Eq. (8.120) in conjunction with Eqs. (8.121)–(8.124):

$$A = \begin{bmatrix} -(1/\tau_\sigma) & (\omega_m - \omega_{sl}) & v_r/(L_\sigma \cdot \tau_r) & v_r \omega_m / L_\sigma \\ -(\omega_m + \omega_{sl}) & -(1/\tau_\sigma) & -[(v_r \omega_m)/L_\sigma] & v_r/(L_\sigma \tau_r) \\ R_r \cdot v_r & 0 & -(1/\tau_r) & \omega_{sl} \\ 0 & R_r \cdot v_r & -(\omega_{sl}) & -(1/\tau_r) \end{bmatrix} \tag{8.128a}$$

The variables used here are defined as follows:

$$v_r = M/L_{rr};\ R_\sigma = R_s + v_r^2 R_r;\ L_\sigma = L_{\sigma s} + v_r L_{\sigma r} = L_{ss} - v_r M;\ \tau_\sigma = L_\sigma / R_\sigma;\ \tau_r = L_{rr}/R_r = (L_{\sigma r} + M)/R_r \tag{8.128b}$$

The input matrix B is given by

$$B = \begin{bmatrix} 1/L_\sigma & 0 \\ 0 & 1/L_\sigma \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{8.129}$$

The output matrix C is given by

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{8.130}$$

It is important to note that the above fourth-order state space model is for induction motors that are powered from voltage source inverters. If instead current source inverter is used, then the stator current components i_{sd} and i_{sq} will become inputs to the system and will no longer be the state variables of the system. Therefore, the only states are the rotor flux linkages ψ_{rd} and ψ_{rq}. As a consequence, if current source inverters are used to supply power to the induction motor, the induction motor torque dynamics will behave as a second-order system.

Stator Flux State Model

In this model, the state vector consists of the stator currents and the stator flux linkages as state variables. Here,

State vector $x = [i_{sd}, i_{sq}, \psi_{sd}, \psi_{sq}]^T$

Input vector $u = [V_{sd}, V_{sq}]^T$

Output vector $y = [i_{sd}, i_{sq}]^T$

The system matrix A is given by

$$A = \begin{bmatrix} -(1/\tau'_\sigma) & \omega_{sl} & (R_r \cdot v_r)/(L_\sigma \cdot M) & \omega_m / L_\sigma \\ -\omega_{sl} & -(1/\tau'_\sigma) & -(\omega_m / L_\sigma) & (R_r \cdot v_r)/(L_\sigma \cdot M) \\ -R_s & 0 & 0 & \omega_a \\ 0 & -R_s & -\omega_a & 0 \end{bmatrix} \tag{8.131}$$

The input matrix B is given by

$$B = \begin{bmatrix} 1/L_\sigma & 0 \\ 0 & 1/L_\sigma \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.132}$$

The output matrix C is given by Eq. (8.130). The variables are defined as follows:

$$R_\sigma' = R_s + R_r(v_r L_{ss}/M);\ t_\sigma' = L_\sigma / R_\sigma'$$

All other variables are as defined in the rotor flux state model.

Air-Gap Flux State Model

In this model, the state vector consists of the stator currents and the air-gap flux linkages as state variables. Here,

State vector $x = [i_{sd}, i_{sq}, \psi_{md}, \psi_{mq}]^T$

Input vector $u = [V_{sd}, V_{sq}]^T$

Output vector $y = [i_{sd}, i_{sq}]^T$

where

$$\psi_{md} = Mi_{sd} + Mi_{rd} \tag{8.133}$$

$$\psi_{mq} = Mi_{sq} + Mi_{rq} \tag{8.134}$$

The system matrix A is given by

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \tag{8.135}$$

where

$$a_{11} = a_{22} = -\frac{R_s + v_r \cdot R_r}{L_\sigma}; a_{33} = a_{44} = -\frac{v_r \cdot L_{\sigma s} \cdot R_r}{L_\sigma \cdot M}$$

$$a_{12} = -a_{21} = \omega_a - \left(\frac{v_r \cdot L_{\sigma r}}{L_\sigma}\right) \cdot \omega_m$$

$$a_{13} = a_{24} = \frac{R_r}{L_\sigma \cdot L_{rr}}; \quad a_{14} = -a_{23} = \frac{\omega_m}{L_\sigma}$$

$$a_{31} = a_{42} = -v_r \cdot \left(\frac{L_{\sigma r}}{L_\sigma}\right) \cdot R_s + v_r \cdot \left(\frac{L_{\sigma s}}{L_\sigma}\right) \cdot R_r$$

$$a_{32} = -a_{41} = \frac{v_r \cdot L_{\sigma s} \cdot L_{\sigma r} \cdot \omega_m}{L_\sigma}$$

$$a_{34} = -a_{43} = \omega_a - \left(\frac{L_{\sigma s}}{L_\sigma}\right) \cdot \omega_m$$

The input matrix B is given by

$$B = \begin{bmatrix} 1/L_\sigma & 0 \\ 0 & 1/L_\sigma \\ (v_r L_{\sigma r})/L_\sigma & 0 \\ 0 & (v_r L_{\sigma r})/L_\sigma \end{bmatrix} \tag{8.136}$$

The output matrix C is given by Eq. (8.130). All variables are as defined in the rotor flux state model of the system. In the next sub-section, a generalized flux state model is conceived which enables switching between any of the above three flux state models dynamically.

Generalized Flux State Model

In speed and torque control applications using induction motor, the stator flux state model is preferred for flux estimation at high speeds because the error due to stator resistance uncertainty becomes negligible at high speeds. But at low speeds, the drop across the stator resistance becomes a significant portion of the applied voltage and therefore the stator flux state model for flux estimation becomes inaccurate. Therefore, for low-speed applications the rotor flux state model is preferred. By appropriate choice of two constants, k_1 and k_2, in the generalized flux state model, the induction motor can be represented in the rotor flux state model, stator flux state model or the air-gap flux state model.

Let a generalized flux linkage in the arbitrary reference frame be defined as follows:

$$\psi_{gd} = M \cdot (k_1 \cdot i_{sd} + k_2 \cdot i_{rd}) \tag{8.137}$$

$$\psi_{gd} = M \cdot (k_1 \cdot i_{sq} + k_2 \cdot i_{rq}) \tag{8.138}$$

If $k_1 = L_{ss}/M$ and $k_2 = 1$, then the flux linkages as defined in Eqs. (8.137) and (8.138) would represent the stator flux linkages given in Eqs. (8.121) and (8.122). If $k_1 = 1$ and $k_2 = L_{rr}/M$, then the flux linkages as defined in Eqs. (8.137) and (8.138) would represent the rotor flux linkages as given in Eqs. (8.123) and (8.124).

Likewise, if $k_1 = 1$ and $k_2 = 1$, then the flux linkages as defined in Eqs. (8.137) and (8.138) would represent the air-gap flux linkages as given in Eqs. (8.133) and (8.134). Thus by a proper choice of the constants k_1 and k_2, the generalized flux linkage definitions of Eqs. (8.137) and (8.138) can be made to represent any given flux linkage in the induction motor.

The rotor currents i_{rd} and i_{rq} can now be represented in terms of the generalized flux linkages as follows:

$$i_{rd} = \frac{\psi_{gd} - M \cdot k_1 \cdot s_{sd}}{M \cdot k_2} \tag{8.139}$$

$$i_{rq} = \frac{\psi_{gd} - M \cdot k_1 \cdot i_{sq}}{M \cdot k_2} \tag{8.140}$$

Substituting for the rotor currents from Eqs. (8.139) and (8.140) into the *d–q* axes motor model of the induction motor as given by Eq. (8.120), one obtains a set of linear differential equations with the stator currents and the generalized flux linkages as the state variables. Here,

State vector $x = [i_{sd}, i_{sq}, \psi_{gd}, \psi_{gq}]^T$

Input vector $u = [V_{sd}, V_{sq}]^T$

Output vector $y = [i_{sd}, i_{sq}]^T$

where ψ_{gd} and ψ_{gq} are defined as per Eqs. (8.137) and (8.138).

The system matrix A is given by

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \tag{8.141}$$

where

$$a_{11} = a_{22} = -\frac{R_s + v_r \cdot v_k \cdot R_r}{L_\sigma}$$

$$a_{33} = a_{44} = -\frac{R_r}{L_{rr}} \cdot \left(1 - \frac{L_{\sigma k} v_r}{L_\sigma}\right)$$

$$a_{12} = -a_{21} = \omega_a - \omega_m \cdot \frac{1}{L_\sigma} \cdot \left(M \cdot v_k - \frac{M^2}{L_{rr}}\right)$$

$$a_{13} = a_{24} = \frac{R_r}{L_\sigma \cdot L_{rr} \cdot k_2}$$

$$a_{14} = -a_{23} = \frac{\omega_m}{L_\sigma \cdot k_2}$$

Table 8.6 Values of the constants k_1 and k_2 for the various flux linkage representation in the generalized flux state motor model

Flux linkage	k_1	k_2	υ_k	$L_{\sigma k}$
Rotor	1	L_{rr}/M	υ_r	0
Stator	L_{ss}/M	1	L_{ss}/M	L_σ/υ_r
Air gap	1	1	1	$L_{\sigma r}$

$$a_{31} = a_{42} = -v_r \cdot k_2 \cdot \left[R_s \cdot \frac{L_{\sigma k}}{L_\sigma} - \left(1 - \frac{v_r \cdot L_{\sigma k}}{L_\sigma} \right) R_r \cdot \upsilon_k \right]$$

$$a_{32} = -a_{41} = -v_r \cdot k_2 \cdot L_{\sigma k} \cdot \omega_m \cdot \left[\frac{1}{L_\sigma} \left(M \cdot v_k - \frac{M^2}{L_{rr}} \right) - 1 \right]$$

$$a_{34} = -a_{43} = \omega_a - \omega_m \cdot \left(1 - \frac{v_r \cdot L_{\sigma k}}{L_\sigma} \right)$$

The input matrix B is given by

$$B = \begin{bmatrix} 1/L_\sigma & 0 \\ 0 & 1/L_\sigma \\ (v_r \cdot k_2 \cdot L_{\sigma k})/L_\sigma & 0 \\ 0 & (v_r \cdot k_2 \cdot L_{\sigma k})/L_\sigma \end{bmatrix} \tag{8.142}$$

The output matrix C is given by Eq. (8.130). The following are the definitions of the variables involved:

$$\nu_k = k_1/k_2;\ L_{\sigma k} = L_{rr}(k_1/k_2) - M$$

All other variables are as defined in the rotor flux state model of the induction motor.

From the generalized flux state model given by Eqs. (8.141) and (8.142), any of the previous models, that is, the rotor flux state model or the stator flux state model or the air-gap flux state model, can be obtained by simply choosing the suitable values of k_1 and k_2 according to Table 8.6.

Electromagnetic Torque Equation

The developed electromagnetic torque T_d in a symmetrical three-phase smooth air-gap induction motor can be expressed in the following general form:

$$T_d = c \cdot [(i_r \cdot e^{j\varepsilon}) \times i_s] \tag{8.143}$$

where under linear magnetic conditions, c is a constant; ε is the rotor position with respect to the stator; $i_r e^{j\varepsilon}$ and i_s are the rotor and stator current space vectors, respectively, referred to the stator reference frame. From the definition of the generalized flux linkages as given in Eqs. (8.137) and (8.138), the *space-vector representation* of the generalized flux linkage space vector is given by

$$\psi_g = M \cdot (k_1 \cdot i_s + k_2 \cdot i_r \cdot e^{j\varepsilon}) \tag{8.144}$$

Using Eq. (8.144) in Eq. (8.143) and using the fact that cross-product of i_s with itself (i.e., $i_s \times i_s$) is 0, the electromagnetic torque of an induction motor is expressed as

$$T_d = c_a \cdot \psi_g \times i_s \tag{8.145}$$

The developed drive torque T_d, expressed in Eq. (8.145), is the cross-product of the generalized flux linkage space vector ψ_g and the stator current space vector i_s; c_a is a constant. By using Table 8.6, the developed torque can be expressed in terms of the rotor flux linkage, stator flux linkage or the air-gap flux linkage as follows:

$$T_d = c_1 \cdot \psi_r \times i_s \tag{8.146}$$

$$T_d = c_2 \cdot \psi_s \times i_s \tag{8.147}$$

$$T_d = c_3 \cdot \psi_m \times i_s \tag{8.148}$$

Thus from Eqs. (8.146)–(8.148), it is evident that the electromagnetic torque can be expressed as the cross-product of the rotor, stator or air-gap flux linkage space vector and the stator current space vector. From this, it is evident that the drive torque is dependent on the sine of the angle between the flux linkage and the stator current space vectors and is independent of their absolute positions. As a consequence, the drive torque expression is *invariant with respect to co-ordinate transformations*. In other words, the evaluation of the drive torque can be performed with respect to any reference frame without loss of generality. With respect to the arbitrary reference frame, the drive torque can be expressed as a function of the resolved components of the generalized flux linkage and stator current space vectors along the d and q axes as follows:

$$T_d = \frac{2}{3} \cdot p \cdot \frac{1}{k_2} \cdot (\psi_{gd} \cdot i_{sq} - \psi_{gq} \cdot i_{sd}) \tag{8.149}$$

The drive torque expressed in Eq. (8.149) is a general expression in terms of the generalized flux linkages. By substituting the proper values of constants k_1 and k_2 from Table 8.6 into Eq. (8.149), the electromagnetic or the drive torque T_d can be expressed in terms of the rotor, stator or air-gap flux linkages as follows:

$$T_d = \frac{2}{3} \cdot p \cdot v_r (\psi_{rd} \cdot i_{sq} - \psi_{rq} \cdot i_{sd}) \tag{8.150}$$

$$T_d = \frac{2}{3} \cdot p \cdot (\psi_{sd} \cdot i_{sq} - \psi_{sq} \cdot i_{sd}) \tag{8.151}$$

$$T_d = \frac{2}{3} \cdot p \cdot (\psi_{md} \cdot i_{sq} - \psi_{mq} \cdot i_{sd}) \tag{8.152}$$

The complete generalized state space description of the induction motor is now given by Eqs. (8.125), (8.126), (8.127), (8.130), (8.141), (8.142), (8.149).

CONCLUDING REMARKS

This chapter brought forth the general principles and methods for obtaining mathematical models for physical systems. This is often the most important and crucial step in understanding and controlling the physical systems. Before attempting to design a controller for a physical system, the model of the physical system should be obtained. The quality of the dynamic control will entirely depend on the accuracy of the mathematical representation of the physical system. The control engineer often has to obtain the mathematical model for a variety of systems with components of diverse nature, such as electrical, mechanical, hydraulic and thermal systems, as well as various combinations of these. In all these cases one must first identify the variables, which may be either flow variables or effort variables. In most cases it is helpful to draw an analogous electrical circuit for which the differential equations are easily obtained by applying Kirchhoff's laws.

One starts by writing the differential equations relating the variables in each component of the system, as well as the interconnection equations. Although in most cases these differential equations are non-linear, as a first approximation they are linearized about the operating points. In this case it is very convenient to identify one variable as the input, or cause, and another variable as output, or effect. This leads to the block diagram for each component, which relates the Laplace transforms of the input and the output through the transfer function. The overall transfer function of the system is then obtained either by simplifying the block diagram or by directly using Mason's rule. However, it should be noted that the transfer function and block diagram approach has one significant drawback and that is the lack of reflection of the loading effect on the input. This is due to the fact that the transfer function and block diagram method are defined for infinite input impedance and zero output impedance. When block diagram algebra is employed with these constraints, the resulting model of the system is inaccurate and far removed from the actual system.

The method of bond graphs is by far the most complete modeling tool that can handle multiple energy domains. Further, the loading effect is inherently reflected upto the first stage, thus giving a model that is much closer to the physical system. The bond graph method also handles non-linear elements in a natural manner. This technique gives an elegant way of obtaining the state equations of physical systems. The circuit averaging method is convenient for the switched-mode power circuits. However, when complex multi-domain systems like motor, pumps, hydraulics, etc. are involved, the bond graph method is most apt.

The space-vector modeling method discussed towards the end of the chapter is a very powerful method especially for polyphase power electronic systems. Modeling AC machines like induction machines, slip ring inductions machines, synchronous machines, three-phase inverter systems, etc. are easily handled by the space-vector methodology.

The tutorial exercises in the next section give few sample practice modeling exercises that should enable one to internalize the modeling process for any given physical system. Many more exercises than provided in the tutorials and the question set should be practiced in order to gain insight into the power electronics system modeling. It should be noted that an accurate and precise model of the physical system will result in a better controller that will be discussed in later chapters.

TUTORIAL EXERCISES

1. Consider the DC–DC converter circuit shown in Figure 8.38. During the dT_s period the transistor Q is ON and during $(1-d)T_s$ period the transistor is OFF and the diode is ON.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Simulation in SciLAB

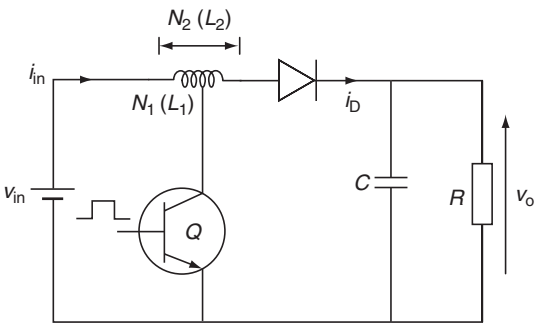


Figure 8.38 | DC–DC converter circuit.

Tasks for study:

(a) Using the circuit averaging method and the bond graph method obtain the averaged large-signal model of the above converter. Verify that the state equation is given as

$$\begin{bmatrix} \frac{di_{in}}{dt} \\ \frac{dv_C}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\left(\frac{N_1}{N_2}\right)\left(\frac{1-d}{L_1}\right) \\ \left(\frac{N_1}{N_2}\right)\left(\frac{1-d}{C}\right) & -\frac{1}{RC} \end{bmatrix} \cdot \begin{bmatrix} i_{in} \\ v_C \end{bmatrix} + \begin{bmatrix} \frac{d}{L_1} + \left(\frac{N_1}{N_2}\right)\left(\frac{1-d}{L_1}\right) \\ 0 \end{bmatrix} \cdot \begin{bmatrix} v_{in} \end{bmatrix}$$

(b) Obtain the steady-state model from the averaged large-signal model. From the steady-state model obtain V_o/V_{in} relationship.

(c) If $V_{in} = 15$ V, $D = 0.3$, N_1:N_2 is 1:2. Evaluate V_o.

(d) Obtain the small-signal model for the above system.

(e) Simulate the DC–DC converter in either spice or SciLab.

(f) Simulate the obtained large-signal and small-signal model equations in SciLAB or MATLAB and compare the state variable waveforms with the circuit waveforms of Step (e).

2. Consider the push–pull converter circuit shown in Figure 8.39. Ideal switches, diodes and transformer can be assumed.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice
b. Simulation in SciLAB

Tasks for study:

(a) Using the circuit averaging method and the bond graph method obtain the averaged large-signal model of the above converter.

(b) Obtain the steady-state model from the averaged large-signal model. From the steady-state model obtain V_o/V_{in} relationship.

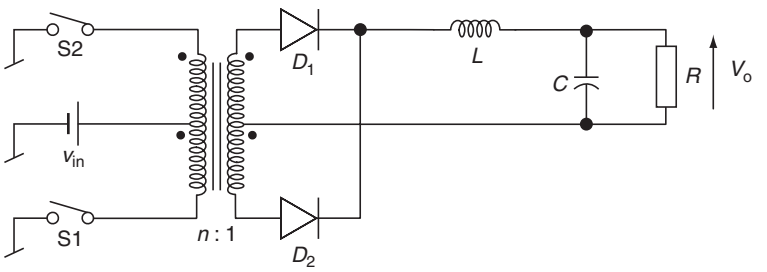


Figure 8.39 | Push–pull converter circuit.

(c) Obtain the small-signal model for the above system. Verify that the small-signal model is given as

$$\begin{bmatrix} \frac{d\hat{i}_L}{dt} \\ \frac{d\hat{v}_C}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\left(\frac{1}{L}\right) \\ \left(\frac{1}{C}\right) & -\frac{1}{RC} \end{bmatrix} \cdot \begin{bmatrix} \hat{i}_L \\ \hat{v}_C \end{bmatrix} + \begin{bmatrix} \left(\frac{D}{n \cdot L}\right) & \left(\frac{V_{in}}{n \cdot L}\right) \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{v}_{in} \\ \hat{d} \end{bmatrix}$$

(d) Simulate the DC–DC converter in either spice or SciLab.

(e) Simulate the obtained large-signal and small-signal model equations in SciLAB or MATLAB and compare the state variable waveforms with the circuit waveforms of Step (d).

3. Consider the non-isolated Cuk converter circuit shown in Figure 8.40. The non-idealities with respect to the inductor and capacitor are included in the form of lumped resistors r_{L1}, r_{L2}, r_{C1} and r_{C2}.

Mode of implementation: The above circuit can be studied by

a. Simulation in Spice

b. Simulation in SciLAB

Tasks for study:

(a) Consider i_{L1}, i_{L2}, v_{C1} and v_{C2} as the state variables.

(b) Using the circuit averaging method and the bond graph method obtain the averaged large-signal model of the above converter.

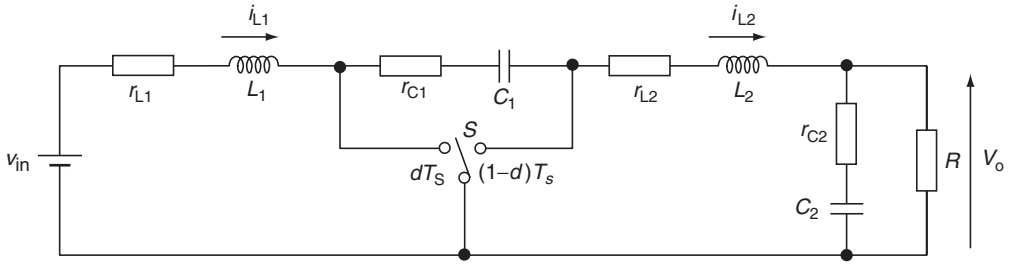


Figure 8.40 | Non-isolated Cuk converter circuit.

(c) For $L_1 = 1.94$ mH, $L_2 = 0.96$ mH, $r_{L1} = 0.17\ \Omega$, $r_{L2} = 0.067\ \Omega$, $C_1 = 850\ \mu$F, $C_2 = 45\ \mu$F, $r_{C1} = 0.05\ \Omega$, $r_{C2} = 0.1\ \Omega$, $R = 30\ \Omega$, $V_{in} = 10$ V and $D = 0.65$, verify that the large-signal model during the DT_s period is

$$\begin{bmatrix} 1.94e-3 & 0 & 0 & 0 \\ 0 & 0.96e-3 & 0 & 0 \\ 0 & 0 & 850e-6 & 0 \\ 0 & 0 & 0 & 45e-6 \end{bmatrix} \cdot \begin{bmatrix} \dot{i}_{L1} \\ \dot{i}_{L2} \\ \dot{v}_{C1} \\ \dot{v}_{C2} \end{bmatrix} = \begin{bmatrix} -0.17 & 0 & 0 & 0 \\ 0 & -0.2166 & -1 & -0.9967 \\ 0 & 1 & 0 & 0 \\ 0 & 0.9967 & 0 & -0.0332 \end{bmatrix} \begin{bmatrix} i_{L1} \\ i_{L2} \\ v_{C1} \\ v_{C2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} v_{in}$$

and for the $(1 - D)T_s$ duration verify that the large-signal model is

$$\begin{bmatrix} 1.94e-3 & 0 & 0 & 0 \\ 0 & 0.96e-3 & 0 & 0 \\ 0 & 0 & 850e-6 & 0 \\ 0 & 0 & 0 & 45e-6 \end{bmatrix} \cdot \begin{bmatrix} \dot{i}_{L1} \\ \dot{i}_{L2} \\ \dot{v}_{C1} \\ \dot{v}_{C2} \end{bmatrix} = \begin{bmatrix} -0.22 & 0 & -1 & 0 \\ 0 & -0.167 & 0 & -0.9967 \\ 1 & 0 & 0 & 0 \\ 0 & 0.9967 & 0 & -0.0332 \end{bmatrix} \begin{bmatrix} i_{L1} \\ i_{L2} \\ v_{C1} \\ v_{C2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} v_{in}$$

(d) Obtain the steady-state model from the averaged large-signal model obtained above. From the steady-state model, obtain V_o/V_{in} relationship.

(e) Obtain the small-signal model for the above system.

(f) Simulate the DC–DC converter in either spice or SciLab.

(g) Simulate the obtained large-signal and small-signal model equations in SciLAB or MATLAB and compare the state variable waveforms with the circuit waveforms of Step (f).

FILL IN THE BLANKS

1. System identification technique is generally used in _____ of system parameters.
2. The physical modeling techniques obtain the mathematical representation of the physical system from _____.
3. The input–output relations for the system are given in a graphical form as _____ characteristic.
4. The input–output transfer curves do not give any information on the _____ of the output to a given system input.
5. The "through" variable may also be described as _____ variable and the "across" variable can also be described as the _____ variable.
6. The equations of equilibrium in different systems are based on _____ principles.
7. The differential equations characterize the _____ of the system.
8. The smallest set of physical quantities called _____ variables completely determines the _____ of the system in the absence of external excitation.

9. The state variables that define the system state are not _____.

10. For a variable to qualify for consideration as a state variable, it should have the important property of _____.

11. The transfer function of the system is defined as the ratio of the Laplace transforms of the _____ to the _____ under _____ initial conditions.

12. The transfer function gives a _____ time _____ model of the physical system.

13. The block diagram representation is undirectional since the _____ cannot produce the _____.

14. The transfer functions are defined between a single _____ and a _____ output.

15. For multiple-input and multiple-output (MIMO) systems one can represent the input–output relationship through a transfer function _____.

16. The order of the denominator polynomial of the transfer function indicates the _____ of the system.

17. For causal systems, the order of the denominator polynomial of the transfer function is _____ than or _____ the order of the numerator polynomial.

18. In the block diagram approach, there is _____ of loading effect on the preceding block.

19. Absence of loading effect in block diagrams implies that each block of the block diagram has _____ input impedance and _____ output impedance.

20. Lagrange's method deals with scalar quantities like _____ and _____ energy.

21. In the case of mechanical systems, the generalized co-ordinates are the independent _____ or the _____ displacements of the different masses of the systems.

22. In electrical networks, the _____ can be considered to form the generalized co-ordinates.

23. The Lagrange method cannot model _____ power systems and the _____ related to such systems like the switched-mode DC–DC converters.

24. In circuit averaging method, the steady-state solution is obtained by equating the rate of change of the state variables to _____.

25. Bond graph is a _____ tool for modeling physical systems.

26. Bond graphs display both _____ and _____ exchanges between components.

27. One-ports are elements exchanging energy with the system through _____ bond only.

28. Two-ports have _____ bonds for energy exchange with the system.

29. The 0-junction is a multi-port where the _____ on all bonds are same and the _____ add up to zero.

30. The 1-junction is a multi-port where the _____ on all bonds are same and the _____ add up to zero.

31. There can be only one bond with a _____ at the 0-junction.

32. There can be only one bond _____ a causal bar at the 1-junction.

33. Systems that are energized by three-or polyphase energy sources can be modeled by using _____.

34. When three-phase sinusoidal currents are fed to a balanced three-phase induction motor, a _____ space vector is generated.

35. The electromagnetic torque of an induction motor is _____ to frame transformation.

DESCRIPTIVE QUESTIONS

1. Write short notes on the following: (a) system identification; (b) physical modeling.
2. Distinguish between linear and non-linear input–output relationships of a physical system.
3. In plotting the input–output transfer characteristics of a system, which variable is considered independent and which is considered the dependent variable?
4. Explain large-signal model, steady-state model and small-signal model of a physical system.
5. Write short notes on the following: (a) system state; (b) state variables; (c) state space; (d) order of the system.
6. What variables qualify to be considered as state variables? Discuss.
7. Discuss on the state equation and output equation of a state space representation of a physical system.
8. Write short notes on the following: (a) linear time-invariant system; (b) linear time-varying system.
9. How does the transfer function method of representing a physical system differ from the state space representation?
10. What are poles and zeros of the systems? Give a physical interpretation for the poles and zeros of a system.
11. Discuss the block diagram representation of systems.
12. Explain the block diagram algebra or block diagram reduction process.
13. Explain Lagrange's method of modeling physical systems.
14. What are the limitations of the Lagrange's method?
15. Explain the circuit averaging method of modeling switched power circuits.
16. What are the merits of the circuit averaging modeling method?
17. Write short notes on the following: (a) large-signal model; (b) averaged large-signal model; (c) steady-state model; (d) small-signal model for switched power circuits.
18. What are the main features of the bond graphs that make it an attractive multi-energy domain modeling method?
19. How are the bond graph element classified?
20. What is one-port? Discuss the one-port bond graph elements and their governing equations.
21. What is two-port? Discuss the two-port bond graph elements and their governing equations.
22. What is multi-port? Discuss the multi-port bond graph elements and their governing equations.
23. Write short notes on the following: (a) effort variable; (b) flow variable in bond graph.
24. What are the power variables for the different energy domains? What is the criterion that needs to be considered to qualify a variable as effort or flow in a particular energy domain?
25. Distinguish between impedance transformer and class transformer.
26. What is causality? List the governing equations of the bond graph elements for flow causality.
27. List the governing equations of the bond graph elements for effort causality.
28. What are the rules for the selection and assignment of causality for a bond graph?
29. What are the steps in the bond graph methodology to obtain the model of a physical system?

30. Why are switched junctions used in bond graph?

31. What are switched 0 and switched 1 junctions? Why are they considered to be generalizations of the 0 and 1 junctions?

32. Discuss the representation of space vector wherein currents are fed to three axes that are place 120° apart in a co-planar space.

33. How are the three-axis components converted to two-axis components? What is the transformation matrix?

34. How are the variables in a stationary reference shifted to an arbitrary rotating reference? Discuss.

35. Discuss the modeling of the induction motor to obtain the *d–q* axis model.

36. Explain the method of obtaining the state space representation of the induction motor from the *d–q* axis model?

37. What are the state variables of the induction motor? What is the order of the system?

38. What is generalized flux state model of the induction motor? What are its features?

39. What are the different representations of the electromagnetic torque in an induction motor?

40. Why is electromagnetic torque of an induction motor invariant with respect to frame transformation?

PROBLEMS

1. Are the following input–output relationships linear or non-linear?

a. $y = au + b$

b. $y = au + by$

c. $y = au^2 + b$

d. $y = a \cdot (u \cdot y) + b$

e. $y = a \cdot \arctan(u) + b$

2. From the small-signal model of the buck converter, obtain the transfer functions for the small-signal deviations of output with respect to the input and duty cycle, that is, $(\hat{v}_o / \hat{v}_{in})(s)$ and $(\hat{v}_o / \hat{d})(s)$.

3. Derive the large-signal, averaged large-signal, steady-state and small-signal models for a boost converter.

4. From the small-signal model of the boost converter, obtain the transfer functions for the small-signal deviations of output with respect to the input and duty cycle, that is, $(\hat{v}_o / \hat{v}_{in})(s)$ and $(\hat{v}_o / \hat{d})(s)$.

5. Boost converters are used as front-end converters wherein the input or inductor current is controlled to be in-phase with the input voltage waveshape. This is discussed in Chapter 10. To achieve input current control, from the small-signal model of the boost converter obtain the transfer function, $(\hat{i}_{in} / \hat{d})(s)$. (Hint: Consider i_{in} as one more output of the output equation. The output matrix of C matrix will be a two row, two column matrix).

6. Derive the large-signal, averaged large-signal, steady-state and small-signal models for a buck–boost converter.

7. From the small-signal model of the buck–boost converter, obtain the transfer functions for the small-signal deviations of output with respect to the input and duty cycle, that is, $(\hat{v}_o / \hat{v}_{in})(s)$ and $(\hat{v}_o / \hat{d})(s)$.

8. Include the non-idealities like the inductor winding resistances, diode drops, switch saturation drops and capacitor equivalent series resistance (ESR) into the dual switch forward

converter circuit. Derive the large-signal, averaged large-signal, steady-state and small-signal models.

9. Include the non-idealities like the winding resistances, diode drops, switch saturation drops and capacitor ESR into the flyback converter circuit. Derive the large-signal, averaged large-signal, steady-state and small-signal models.

10. Given a space vector, how can it be resolved into (a) three-axis components wherein the three axes are 120° apart and (b) two-axis components wherein the two axes are orthogonal to each other.

11. The currents that are applied to the three axes that are 120° apart in space are $i_{sa} = I\sin\omega t$; $i_{sb} = I\sin(\omega t - 2\pi/3)$; $i_{sc} = I\sin(\omega t - 4\pi/3)$ where the frequency of the current waveforms is 50 Hz. Transform these currents into space vectors that are represented in a two orthogonal axes space. What are the current frequencies in the two orthogonal axes if the orthogonal axes are rotating at a frequency of (a) 0 Hz, (b) 25 Hz and (c) 50 Hz?

12. Derive relationships for the following frame transformations:

- **a.** Transform variables from three-phase stationary reference frame to two-phase stationary reference frame.
- **b.** Transform variables from two-phase stationary reference frame to two-phase rotating reference frame rotating at speed ω_a.
- **c.** Transform variables from two-phase rotating reference frame that is rotating at speed ω_a to two-phase stationary reference frame.
- **d.** Transform variables from two-phase stationary reference frame to three-phase stationary reference frame.

ANSWERS

Fill in the Blanks

1. on-line estimation
2. first principles
3. input–output transfer
4. time evolution
5. "flow"; "effort"
6. energy conservation
7. dynamics
8. state; evolution
9. unique
10. memory
11. output; input; zero
12. linear; invariant
13. "effect"; "cause"
14. input; single
15. matrix
16. order
17. greater; equal to
18. absence
19. infinite; zero
20. potential; kinetic
21. linear; angular
22. charges
23. switched; dynamics
24. zero
25. graphical
26. energy; signal
27. one
28. two
29. efforts; flows
30. flows; efforts
31. causal bar
32. without
33. space vectors
34. rotating
35. invariant

CHAPTER 9

Control System Essentials

Learning Objectives

After reading this chapter, you will be able to:

- represent a controller in the digital domain.
- convert analog domain models to discrete domain representations.
- specify performance objectives for a system.

There are some generic issues with respect to control system engineering that are very useful in both analysis and design. Many such aspects like sampling, representation in discrete domain, continuous to discrete domain conversion will be used in many contexts in this and other chapters as well. However, it will not be easy to discuss these issues every time they crop up during the analysis and design in the various chapters. This is therefore a utility chapter that discusses the important general tools and concepts and hence this chapter is titled control system essentials.

Figure 9.1 shows the block diagram of a typical control system. It consists of three major functional blocks, viz. the controller, the plant to be controlled and the feedback block which mainly performs the signal processing on the sensed signal so that it is converted to a compatible form that can be fed back to the controller. Until recently, all the three functional blocks were designed and implemented in the continuous or analog domain. However, in today's systems there is a vertical divide between the controller and the plant to be controlled. Majority of the controllers exist in the digital domain whereas the plant itself exists in the continuous or the analog domain. This means that the feedback signal processing block is functionally subdivided into an analog signal processing (ASP) part and a discrete signal processing (DSP) part.

One may wonder as why such a shift from analog controller to digital controller has occurred even though analog controllers have many advantages like better accuracy, better precision, lesser developmental infrastructure, more compact, better noise immunity, etc. as compared to their digital counterparts. The digital controllers score over the analog controllers in a few significant issues which make them more prevalent in today's control systems. They are as follows

1. **Computation:** Complex control structures involving complex equations can be handled easily.
2. **Repeatability:** Analog circuits have a major problem of parameter drift with temperature which is not an issue in the case of digital circuits. Therefore, in the case of large-scale productions, digital controllers will produce repeatable results.
3. **Reusability:** The same hardware can be used for other applications by re-programming the embedded software for a different application.

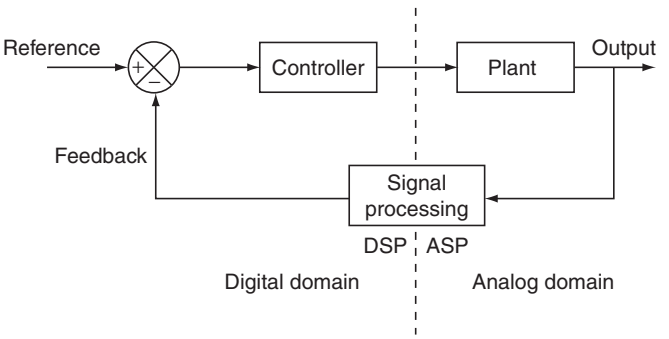


Figure 9.1 General block schematic of a typical control system.

Therefore, most controllers in today's systems will exist in the digital domain. This means that there will be a sampler in the feedback signal processing block that will convert the continuous or analog signals into discrete signals. The transaction between the ASP in the analog domain and the DSP in the digital domain of the signal processing block is through an analog-to-digital converter (ADC). There is one more transaction that will take place between the digital and the analog domains and that is between the controller and the plant. This depends on the nature of the control input that may be given to the plant. If the plant expects a continuous control input, then a digital-to-analog conversion (DAC) is necessary. However, many plants can accept control inputs in the discrete form like pulse-width modulated (PWM) type of signals. In such cases the digital domain output of the controller may be directly interfaced with the discrete control input of the plant without the need of a DAC.

To analyze and design the controller, it is essential that all the three major functional blocks be in a common domain, that is, either all should be equivalently viewed in the digital domain or all should be equivalently viewed in the analog domain. Based on this, one can envisage two approaches to the design of digital controllers:

1. Let the plant and the signal processing functions be expressed in the continuous or the analog domain. Design the controller in the analog domain. Then convert the continuous domain controller to the discrete domain controller.
2. Let the continuous domain plant and the continuous domain signal processing function be converted to the discrete domain. Then design the controller directly in the discrete domain itself.

Now the question arises as to which of the above two approaches is better. In the first approach, the controller is designed in the analog domain and then converted to the discrete domain. This means that the information about the sampling time is absent at the time of controller design. This would imply that the design would work better only when the sampling time tends towards zero. However, low sampling time would put a heavy burden on the digital processor where the controller resides. On the other hand, in the case of the second approach, all the analog domain blocks like the plant and the ASP portion of the signal processing block are first translated into the digital domain models by incorporating the sampling time information. Now the digital controller is designed for the converted discrete plant with the specific sampling time. Therefore, the digital processor can work at higher loop times implying lower overheads and hence a lower cost.

Based on the above discussion, the following control system tools are essential:

1. representing signals and systems in the discrete domain;
2. methods for converting the analog domain models into the discrete domain models.

Apart from the above, the last part of the chapter will discuss few control system basics which will aid the reader in understanding the various control concepts.

9.1 Representation of System in Digital Domain

A system always exists in the time domain. However, the system can be represented in different domains to get more information on the nature and characteristic of the system. Therefore, the physical system can have a mathematical model in every domain in which it can be viewed. In general, systems can be represented in various domains as shown in Table 9.1.

Referring to Table 9.1, the independent variable is "t" in the continuous time domain. In the discrete time domain, the independent variable "t" is discretized at regular intervals of width "T" called the sampling time. In the discrete time domain, the discretized independent variable is called "n". The continuous and the discrete time lines are illustrated in Figure 9.2.

Therefore, the discrete time line, n is given as

$$n = [0T, 1T, 2T, 3T, 4T, \ldots, kT], \; k \to \infty$$

where T is the sampling time. To make the notation simpler, the discrete time line is written as

$$n = [0, 1, 2, 3, 4, \ldots, k]$$

Table 9.1 | Various domains for the signals and systems

Domains	Continuous	Discrete	Used for
Time	t	n	Responses, wave shapes
Frequency	ω	ω	Spectral analysis
Pole-zero	s	z	Analysis, design

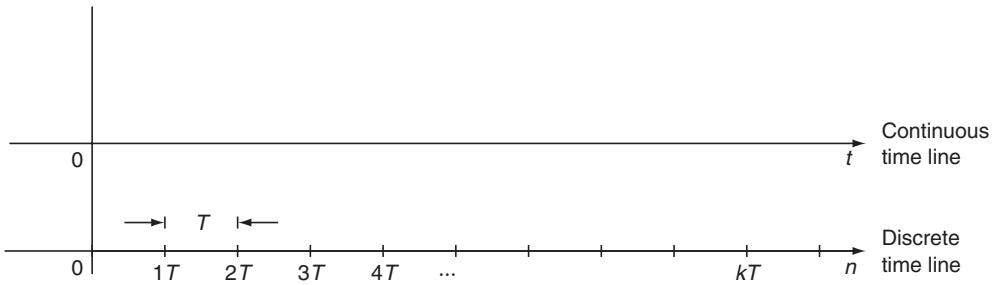


Figure 9.2 | Discretization of the time line.

where it is assumed that the sample intervals are at regular time period T. Thus if one writes the discrete time $n = k$, then it means $n = kT$ units of time.

The frequency domain variable is ω. It is the same for both the continuous and discrete domains. However, in the pole-zero domain, the continuous plane variable is the Laplace variable "s" and the discrete plane variable is the z-transform variable "z". In the pole-zero domain, the continuous plane is called the s-plane and the discrete plane is called the z-plane.

Although a system always exists in the time domain whether it be continuous or discrete, it is important to transit between the various system representations to gather important characteristic information about the system. Further, a difficult analysis in one domain becomes easy in another domain. Therefore, it is essential to transit between domains to increase one's understanding of a system. The system time responses, steady-state accuracies, overshoots, waveshape and their analysis are most appropriately viewed in the time domain. Information like bandwidth, harmonic contents, spectrum, energy spectrum and power spectral density are better understood when viewed in the frequency domain. The pole-zero domain provides information about damping, natural frequencies, time constants and dominant time constants for analysis and design (e.g., about transfer function and location of poles and zeros).

9.2 The *z*-Transform

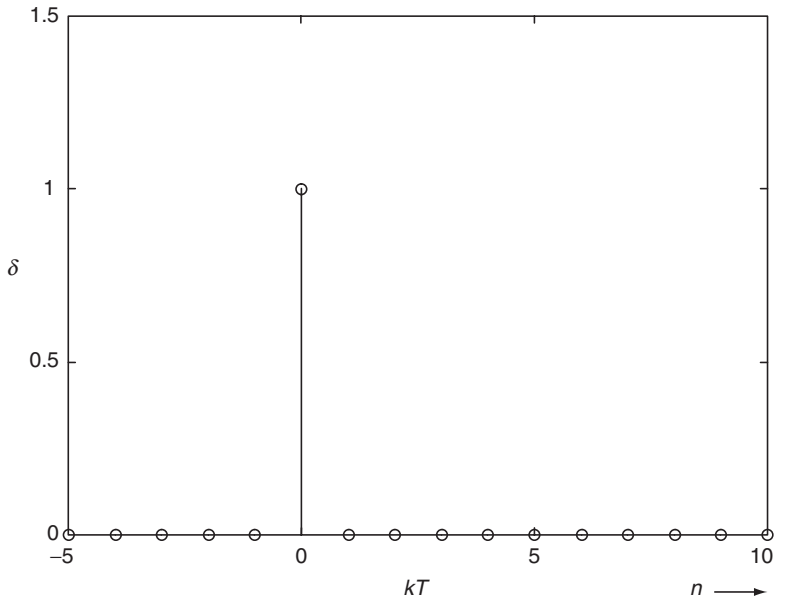
Now consider a continuous signal $x(t)$ which is a function of "t" in the continuous time domain. In any other domain, the same signal is a function of the domain's local variable. The representation of the continuous time signal $x(t)$ in the various domains is indicated in the Table 9.2.

While transiting from an analog domain to a digital domain, two key components are very crucial. One is the analog-to-digital converter itself and the other is the sample-and-hold circuit. The sample-and-hold circuit consists of two important functions: (a) sampling the signal at pre-determined times and (b) holding the value till the next sample so that the ADC can convert the held analog value to the corresponding digital value. The sampling logic signal in practice is of the order of few hundred nanoseconds to a few microseconds to allow for the acquisition of the analog signal by the hold circuit. However, for mathematical representation, the acquisition will be assumed to be instantaneous. Evidently, the impulse signal can be used to represent the sampling instants. Therefore, it is essential to first understand the mathematical representation of a train of impulse which is elucidated by Figures 9.3(a)–(d). Figure 9.3(a) is a single unit impulse occurring at $0T$ point on the discrete time line. This impulse is represented as $\delta(n - 0T)$. If the sampling time T is assumed to exist then it can be represented as $\delta(n - 0)$ or in this case as simply $\delta(n)$.

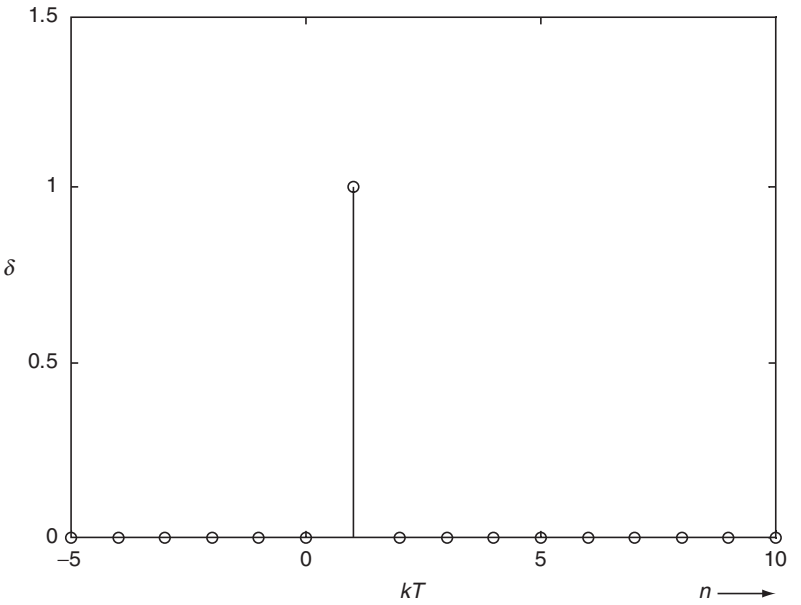
Figure 9.3(b) shows an impulse occurring at the discrete time point of $1T$. This may be viewed as a delayed impulse by $1T$ or it may be viewed as an impulse occurring at $0T$ on a discrete time line which is shifted left

Table 9.2 | Continuous and discrete variables in various domains

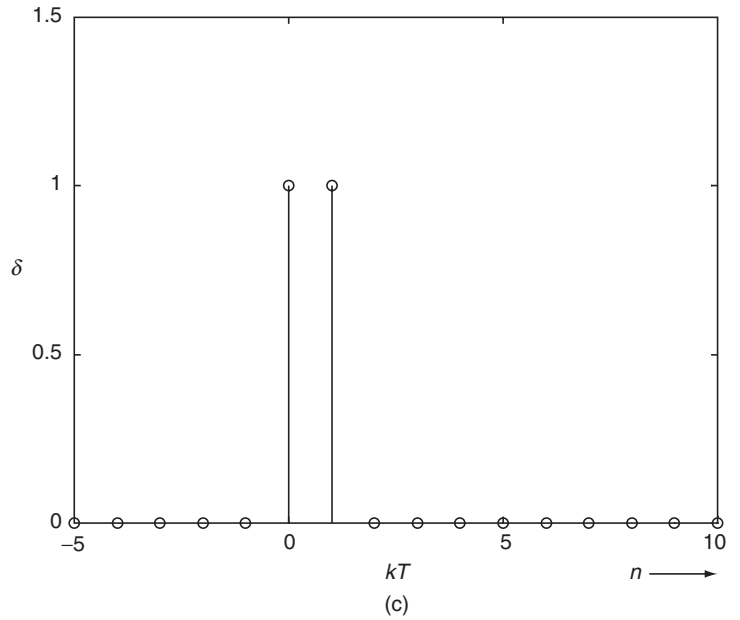
Domains	Continuous	Discrete
Time	$x(t)$	$x(n)$, $x(kT)$ T – sampling time
Frequency	$x(\omega)$	$x(\omega)$
Pole-zero	$x(s)$ (Laplace transform)	$x(z)$ (z-transform)



(a)



(b)



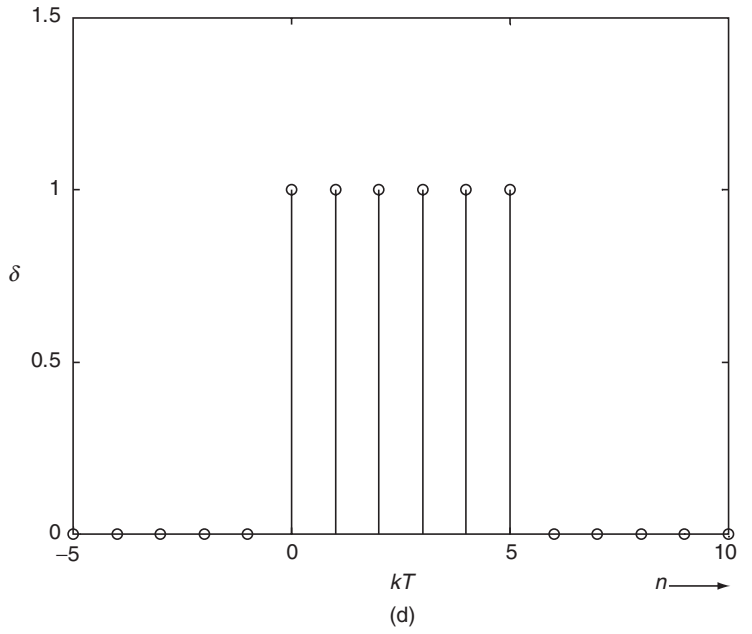


Figure 9.3 (a) Impulse function $\delta(n - 0T)$ or $\delta(n - 0)$ or simply $\delta(n)$; (b) delayed impulse $\delta(n - 1T)$ or simply $\delta(n - 1)$; (c) two impulses $-\delta(n) + \delta(n - 1)$; (d) impulse train $\sum_{k=0}^{5} \delta(n - kT)$ or simply $\sum_{k=0}^{5} \delta(n - kT)$.

by $1T$, that is, $(n-1T)$ time line instead of n time line. In that case, the discrete impulse is represented at $\delta(n-1T)$ or simply $\delta(n-1)$. Figure 9.3(c) shows two impulses one occurring at $0T$ and the other occurring at $1T$ on the discrete time line. Both these impulses together are represented as $\delta(n)+\delta(n-1)$ or $\sum_{k=0}^{1}\delta(n-k)$. Likewise, Figure 9.3(d) is a train of six unit impulses which is represented as $\sum_{k=0}^{5}\delta(n-k)$.

Now, consider the signal $x(t)$ shown in Figure 9.4. For simplicity, the signal $x(t)$ considered is a linear ramp. The signal is sampled with sample interval T. The sample interval is the inverse of the sampling frequency. The sampling frequency is chosen such that it is at least twice the maximum frequency component of the band-limited signal $x(t)$. This means that one has to have a practical knowledge of the frequency spectrum of $x(t)$ to decide the sampling frequency. To discretize the signal, all one needs to do is multiply $x(t)$ or modulate $x(t)$ with a train of unit impulses. This sampled signal is passed through a hold device which holds the value of the impulse sampled value till the next sampling instant. This discretized signal $x(n)$ is shown in Figure 9.4. The discrete domain representation is as follows:

$$\begin{aligned} x(n) &= x(0)\delta(n)+x(1)\delta(n-1)+\cdots \\ &= \sum_{k=0}^{\infty} x(k)\cdot\delta(n-k) \end{aligned} \tag{9.1}$$

where $x(k)$ is the value of $x(t)$ at the sample instant kT. Equation (9.1) is the discrete representation of the continuous domain signal $x(t)$. However, the representation as given in Eq. (9.1) is not very amenable to mathematical analysis. Therefore, to convert the representation of Eq. (9.1) to a more useful formulation that is amenable to mathematical analysis and computer manipulations, the following definition is made: Define

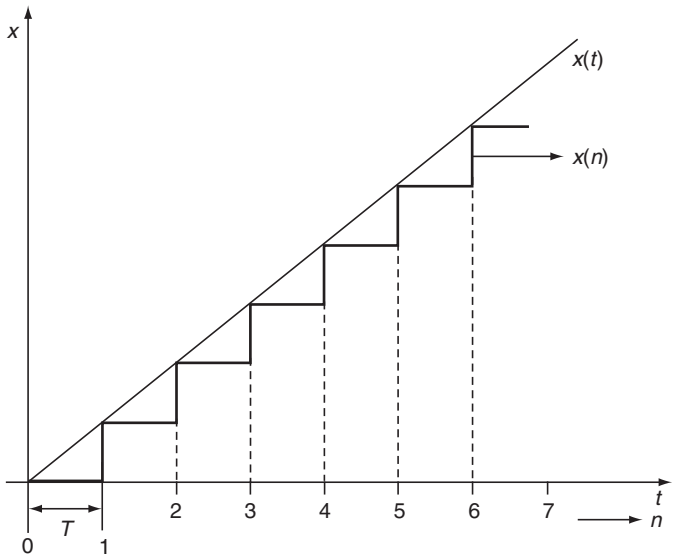


Figure 9.4 | A continuous signal $x(t)$ and its discretized version $x(n)$.

$$X(z) \triangleq \sum_{k=0}^{\infty} x(k) \cdot z^{-k} \tag{9.2}$$

Comparing Eqs. (9.1) and (9.2), one can observe that the impulse component of Eq. (9.1) has been replaced by z^{-k}. $X(z)$ is called the z-transform of $x(n)$. By making a definition as given in Eq. (9.2), a very major advantage has been gained. The discrete representation of Eq. (9.1) is now converted into a polynomial in z^{-1}. This significant accomplishment is achieved just by re-defining the discrete representation $x(n)$ as $X(z)$. Further, one may note that the coefficients of the n-domain representation of the signal as given in Eq. (9.1) and that of the z-transformed representation as given in Eq. (9.2) are the same. Hence there exists a one-to-one correspondence between the signal representation in the n-domain and the z-domain. This means that one may shift between n-domain and z-domain just by inspection as the coefficients are identical in both the domains.

Physical Significance of z^{-1}

If the z-domain representation of $x(n)$ [i.e., Eq. (9.2)] is multiplied by z^{-1}, then

$$X(z) \cdot z^{-1} = \sum_{k=0}^{\infty} x(k) \cdot z^{-(k+1)} \tag{9.3}$$

This when reconstructed will give a signal that is delayed with respect to the original signal $x(t)$ by one unit time step (i.e., $1T$). This is illustrated by the following example.

EXAMPLE 9.1

Consider the signal $x_1(t)$ shown in Figure 9.5. The signal is sampled at regular intervals of time T as indicated in the figure.

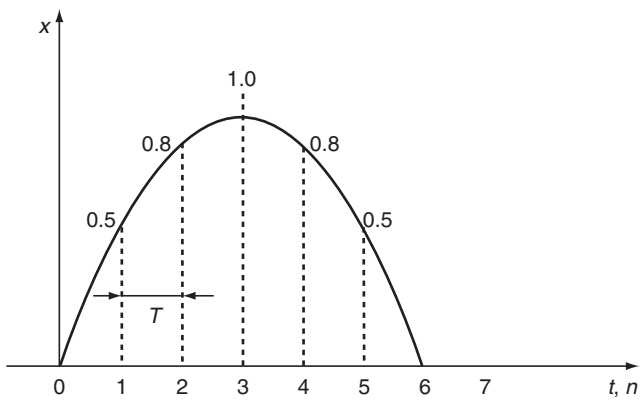


Figure 9.5 | An example signal $x_1(t)$.

Based on Eq. (9.2), the z-transform of $x_1(t)$ can be written by inspecting Figure 9.5 as

$$X_1(z) = 0z^{-0} + 0.5z^{-1} + 0.8z^{-2} + 1z^{-3} + 0.8z^{-4} + 0.5z^{-5} + 0z^{-6} \tag{9.4}$$

Now, multiply both sides of Eq. (9.4) by z^{-1}. One obtains

$$X_2(z) = z^{-1} \cdot X_1(z) = 0z^{-1} + 0.5z^{-2} + 0.8z^{-3} + 1z^{-4} + 0.8z^{-5} + 0.5z^{-6} + 0z^{-7} \quad (9.5)$$

Referring to Eq. (9.5), and by inspecting the coefficients of the power of z^{-1}, one can re-construct the n-domain signal $x_2(t)$. This is shown in Figure 9.6.

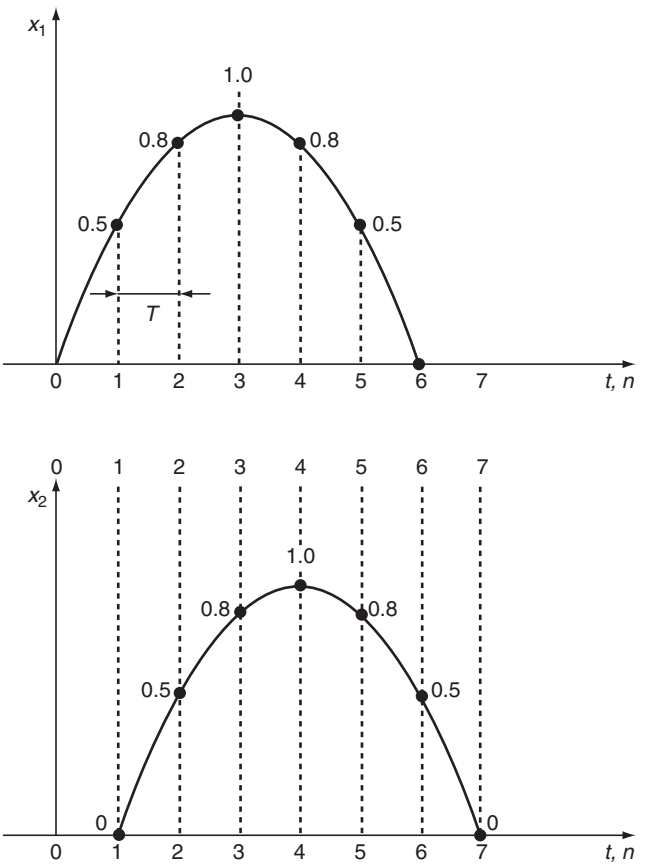


Figure 9.6 | Re-construction of $x_2(t)$ from $X_2(z)$.

It is clear from Figure 9.6 that $x_2(t)$ is a unit time delayed version of $x_1(t)$. This implies that passing a signal through a z^{-1} block introduces a unit time delay. If one were to pass a signal through a z^{-m} block, then the input signal will be time-shifted by m time delays or by mT. While implementing a unit delay in hardware, one will have to pass the input signal in a memory element for a period of $1T$. Then after $1T$ duration has elapsed, the value in the memory is released as output and the current input is now stored in the memory. This way the values that are stored in the memory element and released are delayed by $1T$. If one needs to delay the input signal by $2T$, then the output of the memory element is passed through another memory element to obtain a further $1T$ delay, so that the over all delay is $2T$. This means that for every delay an

additional memory element is required. Therefore, in a physical sense, z^{-1} is nothing by one memory element in the digital processor and z^{-m} means that there are m memory elements in the digital processor. This implies that if there are m data variables, then m memory locations are needed and the order of the system is m as each memory location is a state variable. As a memory element is the basic building block of a digital system, z^{-1} is the basic building block for modeling a discrete system. With just the knowledge of the concept of z^{-1}, any discrete system can be modeled.

9.3 Digital Filter

All systems in the digital domain are considered as digital filters as they modify the waveshape of the signal and hence the harmonic amplitude. Therefore, any design in the digital domain is a digital filter design. There are broadly two distinct types of filters: (a) finite impulse response (FIR) filter and (b) infinite impulse response (IIR) filter.

Consider the discrete system shown in Figure 9.7. This is built by using only z^{-1} blocks and gains blocks. It is called a finite impulse response (FIR) filter.

In Figure 9.7, $x(n)$ is the n-domain input signal and $y(n)$ is the n-domain output signal. The input signal $x(n)$ is passed through a series of unit delays and some gains. The values b_0, b_1, …, b_m are the scalar gains. The output is given as

$$Y(z) = (b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m}) X(z)$$
$$H(z) = Y(z)/X(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m} \tag{9.6}$$

Equation (9.6) gives the transfer function for the FIR filter. $H(z)$ can also be written as

$$H(z) = \frac{b_0 z^m + b_1 z^{m-1} + \cdots + b_m}{z^m} \tag{9.7}$$

Equation (9.7) is the transfer function of an FIR filter of order m. The order of filter is the number of unit delay element needed or the number of poles. From Eq. (9.7) it is evident that for FIR filter all the poles are at $z = 0$. This implies that only the location of the zeros can be designed by the choice of the scalar gains

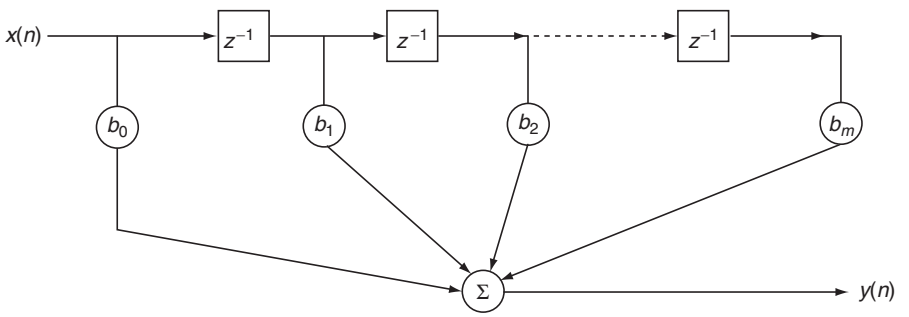


Figure 9.7 | FIR filter.

$b_0, b_1, \ldots, b_m$. Furthermore, it may be seen from Figure 9.7 that an FIR filter is an open-loop filter. Consider a first-order FIR filter. Referring to Eq. (9.6), the transfer function is given as

$$H(z) = b_0 + b_1 z^{-1}$$

If the input is a unit impulse occurring at $0T$, then the output is as shown in Figure 9.8. At time $0T$, a unit impulse is applied to the input of the FIR filter. The output at $0T$ is the unit impulse scaled by b_0. At time $1T$ the unit impulse has shifted to the b_1 line. The output at $1T$ is b_1. Beyond $1T$, the output is zero. There are two important points to be noted. One is that the output response will have the same coefficients as the transfer function because $H(z)$ which is in the z-domain will become $h(n)$ in the n-domain which is nothing but the unit impulse response. Second, the output response will exist upto mT time where m is the order of the FIR filter. This is the reason this type of filter is called finite impulse response filter.

Now, include feedback to the FIR filter shown in Figure 9.7. This is shown in Figure 9.9. The output $y(n)$ is fed back to a series of unit delays. The outputs of each unit delays are tapped and scaled by scalar gains $a_1, a_2, \ldots, a_n$ as indicated in Figure 9.9. After introducing the feedback, the input–output relation is

$$Y(z) = (b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m})X(z) + (a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n})Y(z)$$

Now, the transfer function can be written as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 - a_1 z^{-1} - \cdots - a_n z^{-n}} \tag{9.8}$$

Equation (9.8) gives the transfer function of the filter shown in Figure 9.9. This filter is called an IIR filter as the output will asymptotically die down to zero for a unit input impulse at time $0T$. It should be noted that an FIR filter is a special case of an IIR filter wherein the feedback scalar gains a_0 to a_n are all zero. The number of poles is always greater than or equal to the number of zeros for a causal system. It is also evident from Eq. (9.8) that in the case of an IIR filter both the poles and the zeros can be designed leading to much better cut-off filters. The output response for a unit input impulse is shown in Figure 9.10 which shows that the output response dies down asymptotically to zero as time tends to infinity.

The order of the filter in the case of the FIR filter was pretty straightforward. The number of unit delay elements indicated the order of the filter. However, in the case of an IIR filter the question arises as to what is the order of the filter. Is it the total number of unit delay elements as indicated in Figure 9.9? Or is it the

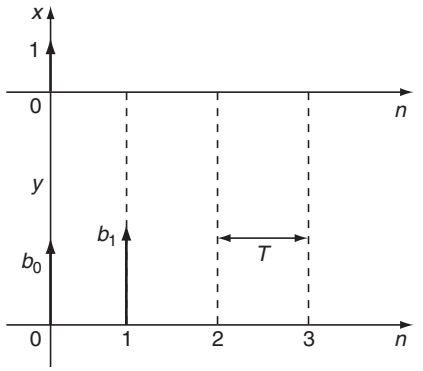


Figure 9.8 | Unit impulse response for FIR filter.

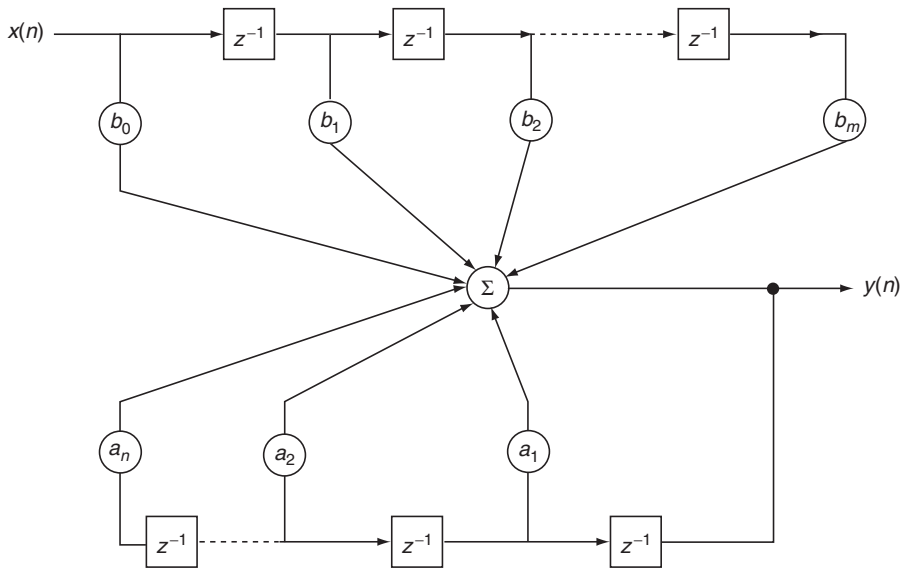


Figure 9.9 | IIR filter.

number of unit delay elements in the forward path? Or is it the number of unit delay elements in the feedback path? To answer this question, let us consider a simple IIR filter with one unit delay element in the forward path and one delay element in the feedback path as shown in Figure 9.11.

Figure 9.11 is drawn in a manner to explicitly indicate the forward and the feedback paths of the filter. The IIR filter shown in Figure 9.11 is divided into two parts. The portion between nodes 1 and 2 is the feed-forward portion and is basically an FIR filter. It has a transfer function

$$\frac{V(z)}{X(z)} = b_0 + b_1 z^{-1}$$

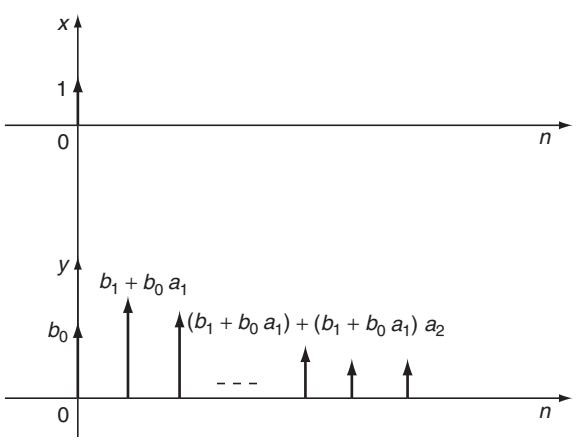


Figure 9.10 | Impulse response of an IIR filter.

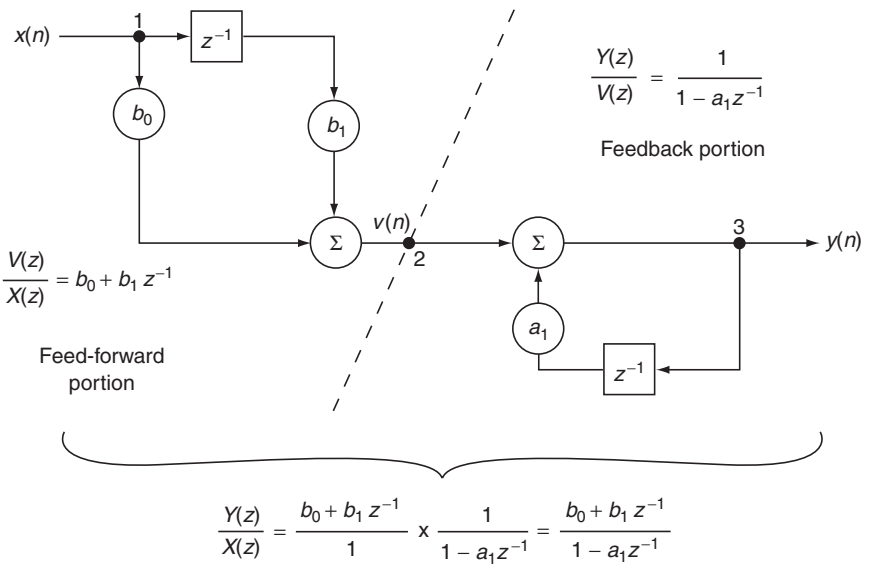


Figure 9.11 | IIR filter with one delay each in feed-forward and feedback paths.

where $v(n)$ is the signal at node 2. The portion between nodes 2 and 3 is the feedback portion of the IIR filter. It has a transfer function given by

$$\frac{Y(z)}{V(z)} = \frac{1}{1 - a_1 z^{-1}}$$

The overall transfer function is given as

$$\frac{Y(z)}{X(z)} = \frac{V(z)}{X(z)} \cdot \frac{Y(z)}{V(z)} = \left(\frac{b_0 + b_1 z^{-1}}{1}\right) \cdot \left(\frac{1}{1 - a_1 z^{-1}}\right)$$

Without loss of generality, the sub-transfer functions could be interchanged as

$$\frac{Y(z)}{X(z)} = \frac{Y(z)}{V(z)} \cdot \frac{V(z)}{X(z)} = \left(\frac{1}{1 - a_1 z^{-1}}\right) \cdot \left(\frac{b_0 + b_1 z^{-1}}{1}\right)$$

On interchanging the sub-transfer functions, the overall transfer function is unchanged. Thereby in the filter block also, the two parts could be interchanged without affecting the overall transfer function. This means the portion of the filter between nodes 2 and 3 will be drawn first and the portion of the filter between nodes 1 and 2 will be drawn next as shown in Figure 9.12. Here nodes 3 and 1 are joined together and the signal is denoted by $w(n)$. Here also it is evident that the overall transfer function is unchanged.

The filter shown in Figure 9.12 can be further simplified as shown in Figure 9.13. In Figure 9.13, it is evident that one of the unit delay elements is redundant because the unit delayed $w(n)$ signal is fed to both the a_1 scalar gain and also to the b_1 scalar gain. Therefore, the unit delay element can be made common to both the arms as indicated in the further simplified diagram Figure 9.13. Thus one sees that there is need for

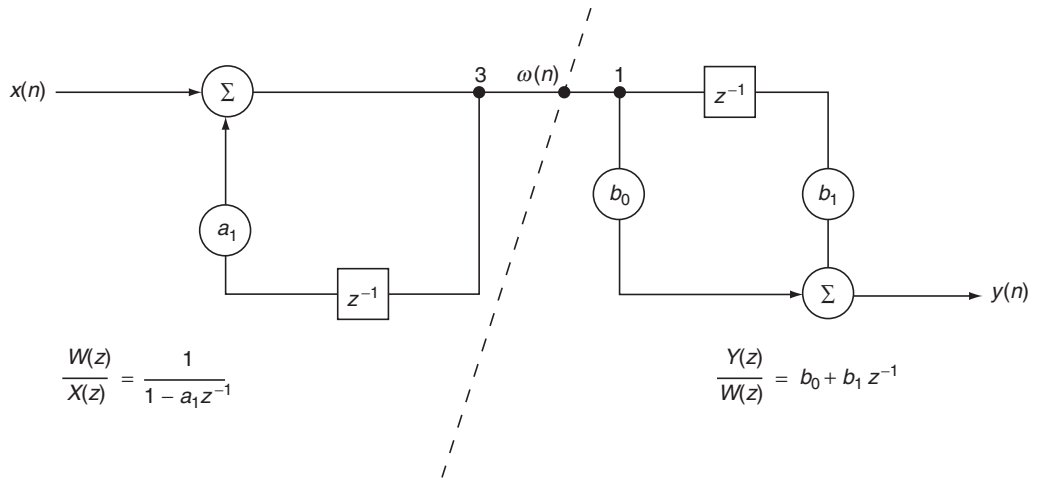


Figure 9.12 | IIR filter with feed-forward and feedback portions interchanged.

only one delay element and therefore the order of the system is one. This simplified form of the IIR filter is called the standard form or the canonical form. It can be extended for any order by appending to the first-order filter by the dotted portions. For a causal system the number of poles is always greater than or equal to the number of zeros. Therefore, the number of delay elements required to construct the filter without any redundant unit delay elements is the same as the number of poles. Hence one can say that in the case of an IIR filter, the order of the system is the number of poles of the IIR filter.

Few Common Digital Filters

Integrator

Let $b_0 = 0$

$b_1 = 1$

$a_1 = 1$

All other scalar gains of Figure 9.9 are zero. With these scalar gain values, the IIR filter reduces to an Euler's forward integrator topology, the block schematic of which is given in Figure 9.14. From Eq. (9.8), the transfer function is given as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{z^{-1}}{1 - z^{-1}} = \frac{1}{z - 1}$$

Low-Pass Filter

Let $b_0 = 0$

$b_1 = 1 - e^{-aT}$

$a_1 = e^{-aT}$

where T is the sampling time. All other scalar gains of IIR filter as shown in Figure 9.9 are zero. With the above values of the scalar gains, one obtains a low-pass filter that is equivalent to $1/(s + a)$. The block schematic of the low-pass filter is shown in Figure 9.15.

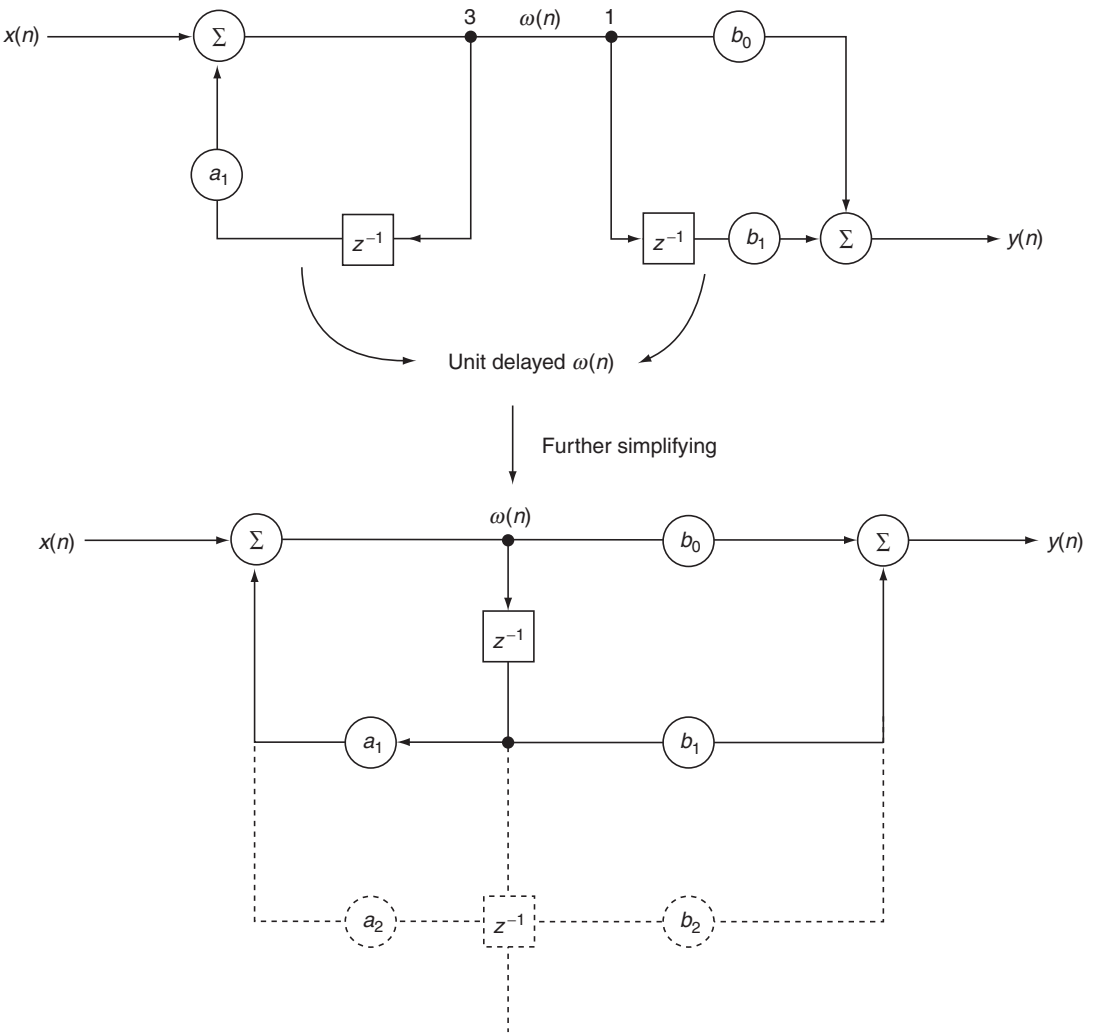


Figure 9.13 | Canonical form of the IIR filter.

From Eq. (9.8), the transfer function is given as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{(1 - e^{-aT})z^{-1}}{1 - e^{-aT}z^{-1}} = \frac{1 - e^{-aT}}{z - e^{-aT}}$$

Differentiator

To implement a first-order differentiator, set

$b_0 = 1/T$

$b_1 = -1/T$

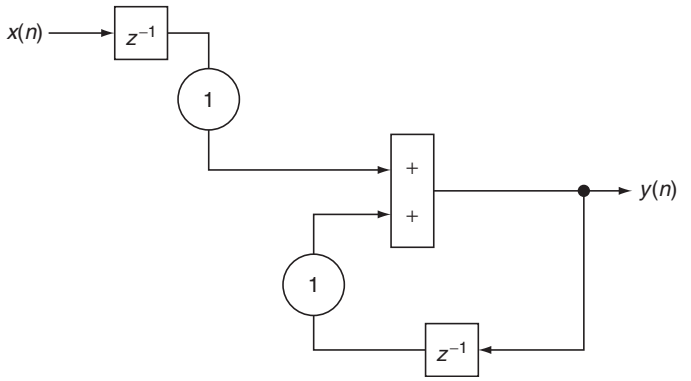


Figure 9.14 | Block schematic of the integrator.

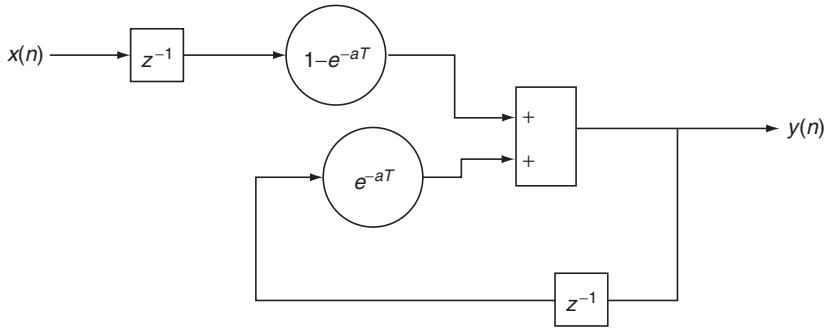


Figure 9.15 | Block schematic of low-pass filter.

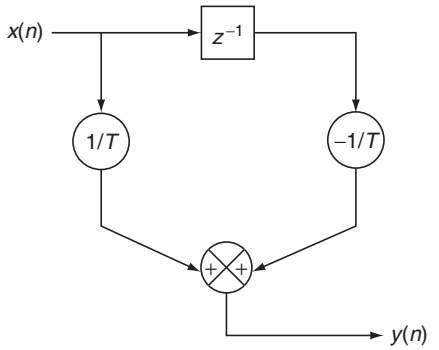


Figure 9.16 | Block schematic of first-order differentiator.

All other scalar gains of the FIR filter of Figure 9.7 are zero. It should be noted that the differentiator is an FIR filter that does not have any feedback terms. The schematic of the first-order differentiator is shown in Figure 9.16.

From Eq. (9.6), the transfer function is given as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{T} + \frac{1}{T} z^{-1} = \frac{1}{T}\left(\frac{z+1}{z}\right)$$

9.4 Mapping between *s*-Plane and *z*-Plane

The physical system exists in the continuous or analog domain. When expressed in the digital domain with the sampling information incorporated, it would give a better understanding if one is able to relate critical *s*-domain parameter in the *z*-domain. The graphical visualization of the relation between *s*-plane and *z*-plane is discussed in this section.

The *z*-plane variable "*z*", the *s*-plane variable "*s*" and the sampling time T are related by the following exponential relation:

$$z = e^{sT} \tag{9.9}$$

For $s = \sigma + j\omega$, from Eq. (9.9) one obtains

$$z = e^{(\sigma + j\omega)T} = e^{\sigma} e^{j\omega T} \tag{9.10}$$

Equation (9.10) is of the form $re^{j\theta}$ in polar co-ordinates wherein

$$r = e^{\sigma} \text{ and } \theta = \omega T$$

It should be noted that the Cartesian co-ordinate system is used in *s*-plane and the polar coordinate system is used in *z*-plane on account of the mapping rule.

$s = 0 + j0 \rightarrow z = 1$

Consider the origin in the *s*-plane wherein $s = 0 + j0$. Substituting for the *s*-variable in Eq. (9.10) one obtains

$$z = e^{0} e^{j0T} = 1$$

This means that the origin of the *s*-plane maps to $z = 1$ point in the *z*-plane as shown in Figure 9.17.

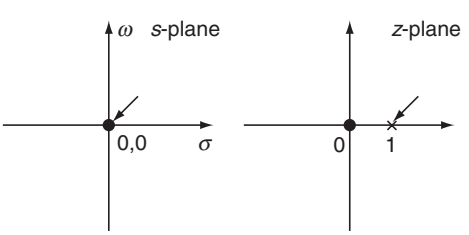


Figure 9.17 | Mapping of $s = 0 + j0$ point in *z*-plane.

$s = 0 + j\pi/T \rightarrow z = -1$

Consider the point $s = 0 + j\pi/T$ in the s-plane. By using Eq. (9.10), one obtains

$$z = e^{0} e^{j\pi} = 1 \cdot e^{j\pi} = -1$$

This mapping is shown in Figure 9.18(a).

Consider the points all along the ω-axis in the s-plane upto π/T. The s-plane variable is

$$s = 0 + ja\pi/T, \quad a \leq 1$$

This maps in the z-plane as

$$z = 1 \cdot e^{ja\pi}$$

This is the locus of points whose vectors have an amplitude of one and angle of $a\pi$. For values of a between 0 and 1, one would trace a semicircle (in anti-clockwise direction) in the z-plane as shown in Figure 9.18(b).

$s = 0 - j\pi/T \rightarrow z = -1$

Consider the point $s = 0 - j\pi/T$ in the s-plane. By using Eq. (9.10), one obtains

$$z = e^{0} e^{-j\pi} = 1 \cdot e^{-j\pi} = -1$$

This also maps to the –1 point as indicated in Figure 9.19(a).

Consider the points all along the ω-axis in the s-plane upto $-\pi/T$. The s-plane variable is

$$s = 0 - ja\pi/T, \quad a \leq 1$$

This maps in the z-plane as

$$z = 1 \cdot e^{-ja\pi}$$

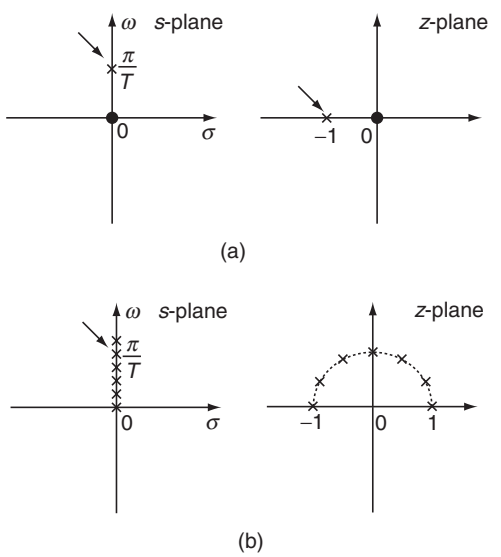


Figure 9.18 | Mapping of (a) $s = 0 + j\pi/T$; (b) $s = 0 + ja\pi/T$.

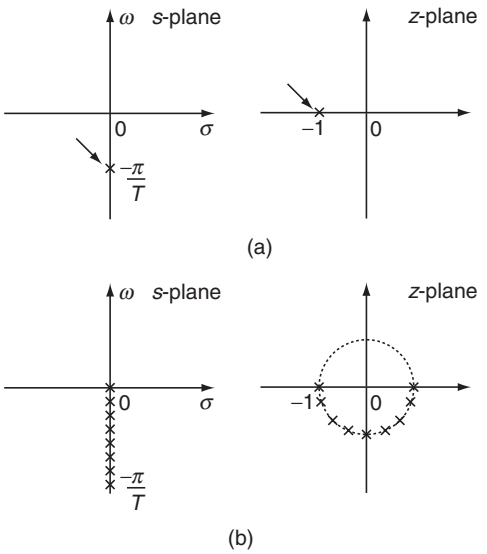


Figure 9.19 | Mapping of (a) $s = 0 - j\pi/T$; (b) $s = 0 - ja\pi/T$.

This is the locus of points whose vectors have an amplitude of 1 and angle of $-a\pi$. For values of a between 0 and 1, one would trace a semicircle (in clockwise direction) in the z-plane as shown in Figure 9.19(b). It should be noted that every $2\pi/T$ maps onto the same unit circle in the z-plane. This is the consequence of sampling.

$s = -\sigma + j0 \rightarrow z =$ Real Axis between 0 and 1

For s = negative real axis, the z-plane mapping is given by

$$z = e^{-\sigma} e^{j0} = \frac{1}{e^{\sigma}}$$

From the above equation it is evident that $\sigma = 0$ maps to $z = 1$ and $\sigma \rightarrow \infty$ maps to $z = 0$. Therefore, the whole of the negative real axis maps to the real axis between 0 and 1 in the z-plane.

Consider the points $s = -\sigma + j\pi/T$. The z-plane mapping is given by

$$z = e^{-\sigma} e^{j\pi} = \frac{-1}{e^{\sigma}}$$

Here, $s = 0 + j\pi/T$ maps to $z = -1$ and $s = -\infty + j\pi/T$ maps to 0. It can be seen from Figure 9.20 that it maps to the real axis between 0 and -1.

The entire left half of the s-plane between $\omega = -\pi/T$ and $\omega = \pi/T$ maps within the unit circle in the z-plane. This region is called the stable region. This implies that the stable region in the s-plane maps to the area within the unit circle implying that the points within the unit circle in the z-plane are the stable region.

$s = \sigma + j0 \rightarrow z =$ Real Axis between 1 and Infinity

For s = positive real axis, the z-plane mapping is given by

$$z = e^{\sigma} e^{j0} = e^{\sigma}$$

From the above equation it is evident that $\sigma = 0$ maps to $z = 1$ and $\sigma \rightarrow \infty$ maps to $z \rightarrow \infty$. Therefore, the whole of the positive real axis maps to the real axis between 1 and infinity in the z-plane. The right half

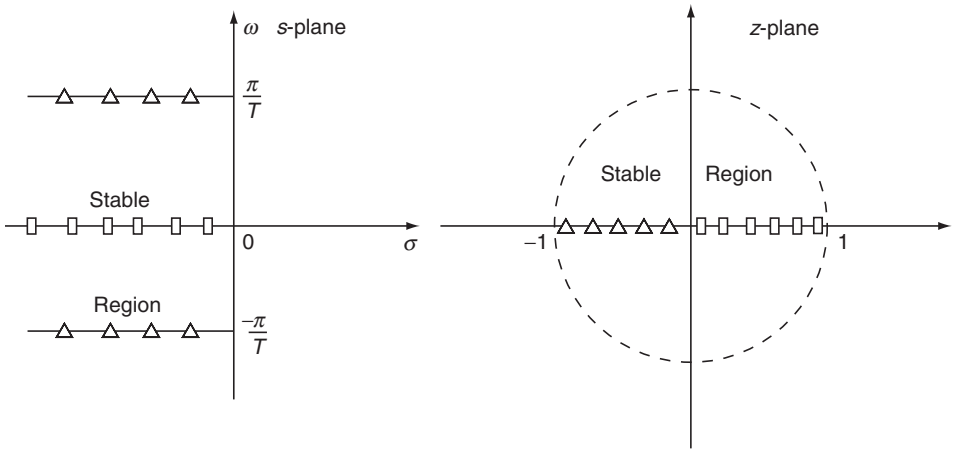


Figure 9.20 | Stable region mapping.

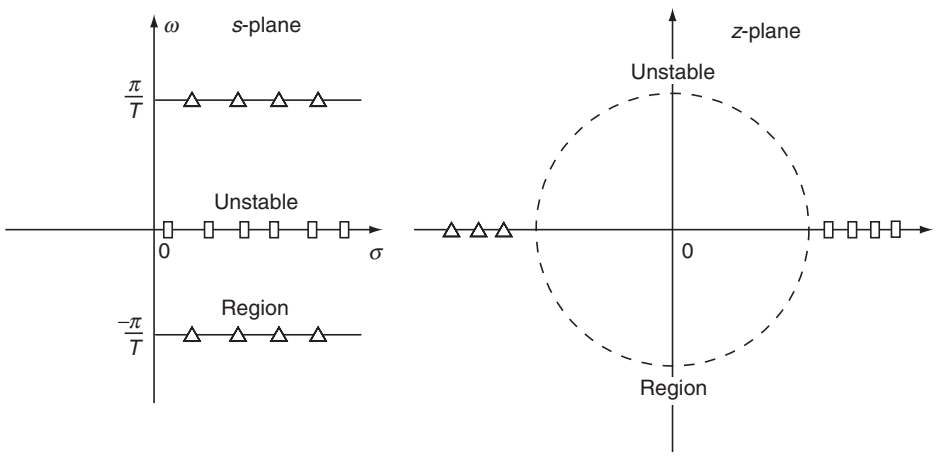


Figure 9.21 | Unstable region mapping.

of the *s*-plane is called the unstable region. By similar arguments it can be shown that the area outside the unit circle in the *z*-plane is the unstable region as indicated in Figure 9.21.

Thus for a discrete system transfer function if the poles and zeros lie inside the unit circle, the system is stable whereas if they are outside the unit circle, the system is unstable.

9.5 Effect of Sampling

A continuous time signal can be converted into a discrete time signal by sampling at a proper sampling frequency. The main question that emerges is with regard to the choice of the sampling frequency and the effect that the sampling will have on the signal.

According to the Nyquist–Shannon sampling theorem, when sampling an analog signal the sampling frequency must be greater than twice the highest frequency component of the analog signal so that one is able to reconstruct the original signal from the sampled version. It should be noted that reconstruction implies that the original waveshape is recoverable. The original amplitude and the phase may not be recoverable.

Consider a continuous signal $x(t)$. Let this signal be sampled at regular intervals of time T called the sampling time. The effect on the various view points of the signal in the various domains is illustrated in Figures 9.22–9.25.

Consider a signal $x(t)$ that has the frequency spectrum as shown in Figure 9.26.

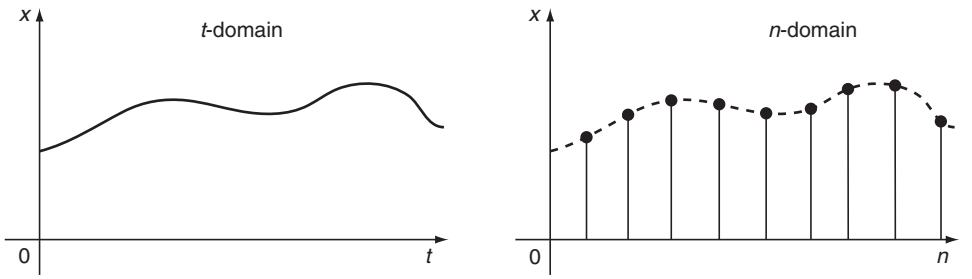


Figure 9.22 | Effect of sampling in time domain.

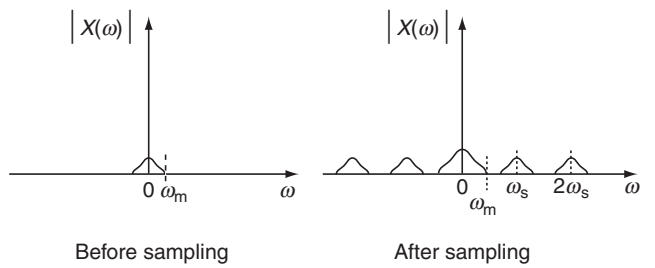


Figure 9.23 | Effect of sampling as seen in the frequency domain.

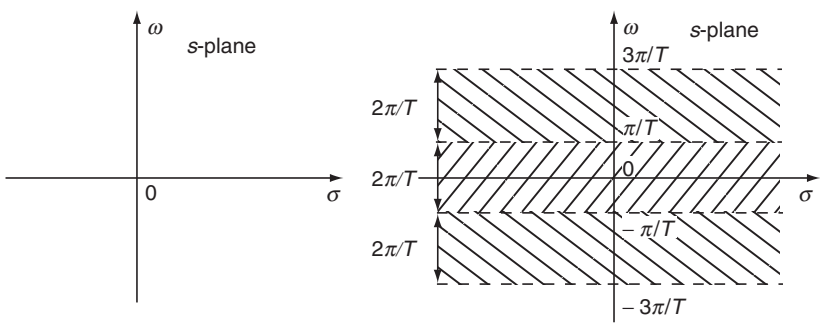


Figure 9.24 | Effect of sampling in the pole-zero domain in the *s*-plane.

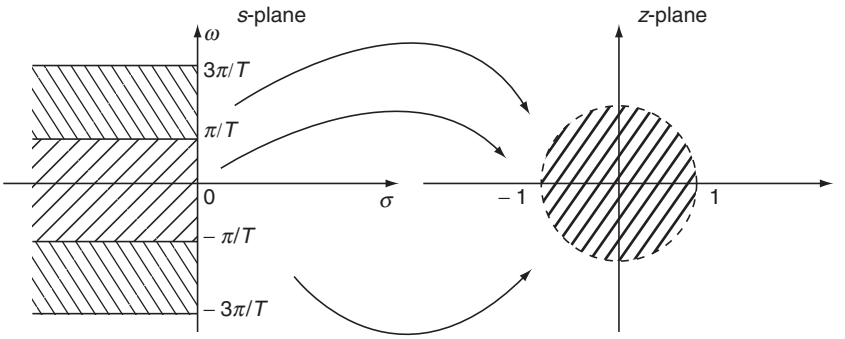


Figure 9.25 | Sampled *s*-plane mapped on to *z*-plane.

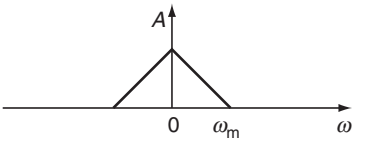


Figure 9.26 | Frequency spectrum of an arbitrary signal $x(t)$.

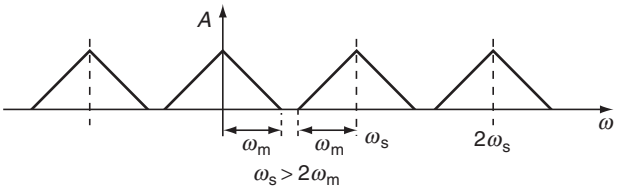


Figure 9.27 | Frequency spectrum of sampled signal $x(t)$ that is sampled with $\omega_s > 2\omega_m$.

If the Nyquist or Shannon sampling criteria is satisfied during sampling then the frequency spectrum of the sampled signal is as shown in Figure 9.27.

However, if the Nyquist criteria are not satisfied during sampling then the frequency spectrum of the under-sampled signal may look like that shown in Figure 9.28. It shows that the sampled spectrums overlap. As a consequence, at the frequencies where there is an overlap one does not know whether one has to take the amplitude corresponding to the high-frequency component of the lower sampled spectrum or the low-frequency component of the higher sampled spectrum. This is called aliasing. Once aliasing occurs it is not possible to reconstruct the signal at all as the amplitude values are not unique in the overlapping frequency regions.

If the sampling is done according to the Nyquist sampling theorem, then the frequency spectrum is obtained as shown already in Figure 9.27. This does not have the aliasing problem. Therefore if one designs a low-pass filter called reconstruction filter as shown in Figure 9.29, then all the higher sampled spectrums will be filtered out and only the fundamental spectrum will remain which is the spectrum of the original signal.

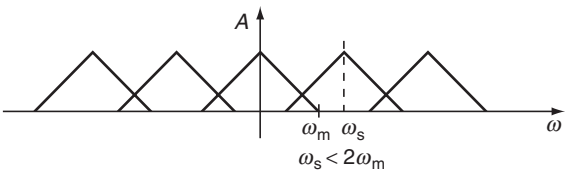


Figure 9.28 | Frequency spectrum of sampled signal $x(t)$ that is sampled with $\omega_s < 2\omega_m$.

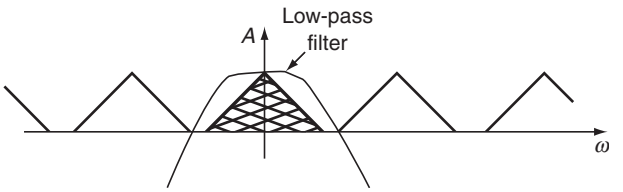


Figure 9.29 | Low-pass filter used for extracting or reconstructing the original signal $x(t)$.

In practice, one does not come across a continuous domain signal $x(t)$ that is so well band-limited as shown in Figure 9.26. Normally due to the presence of noise and unmodeled high-frequency dynamics the actual band-limit may be much higher. Therefore it will become difficult to know the actual value of ω_m and as a consequence it may not be easy to choose the sampling frequency, ω_s. To overcome this problem, the continuous domain signal is passed through a filter, the characteristics of which will be such that it will remove all frequency components beyond a defined frequency value. Thus the output of the filter will have a well-defined band-limit. This band-limited signal when sampled at a higher rate than the Nyquist rate will result in proper sampled signal that can be reconstructed. Such a filter that forces a band-limit on the continuous signal is called the anti-alias filter. Therefore it is recommended that all continuous signals be passed through an anti-alias filter so that the original continuous signal becomes band-limited before passing through the sampler.

9.6 Continuous to Discrete Domain Conversion

It was already discussed in the introduction section of this chapter that one needs to convert the continuous domain model of the system to the discrete domain. In doing so, the sampling information is also carried along into the discrete domain model. Therefore, if this discrete domain model which contains the sampling information is used for the analysis and synthesis of the digital controller, the results obtained will be closer to reality. This section will discuss the various methods for converting the continuous domain models to discrete domain models. The conversion methods are:

1. pole–zero map;
2. numerical integration methods;
3. hold equivalence;
4. discretizing state equation.

Pole–Zero Map

Consider a system that has a continuous domain model expressed as a transfer function $H(s)$. It is now desired that $H(s)$ be converted into an equivalent model in the discrete domain as $H(z)$. The conversion is done by applying Eq. (9.9) which is repeated here for clarity.

$$z = e^{sT}$$

The following procedure is used to obtain the discrete transfer function from the continuous transfer function by using the pole-zero map given by Eq. (9.9).

Consider a system that has a continuous domain transfer function model given as

$$H(s) = K\frac{(s+a_1)(s+a_2)\cdots}{(s+b_1)(s+b_2)\cdots} \tag{9.10}$$

Then

1. All the poles of $H(s)$ are mapped according to $z = e^{sT}$. A pole at $s = -a_1$ in the s-plane is mapped to the pole $z = e^{-a_1T}$ in the z-plane.
2. All finite zeros of $H(s)$ are also mapped according to $z = e^{sT}$. A zero at $s = -b_1$ in the s-plane is mapped to the zero $z = e^{-b_1T}$ in the z-plane.
3. If the transfer function $H(s)$ contains less number of finite zeros than the poles, that is, the order of the numerator is less than the order of the denominator, then it implies that the number of zeros amounting to the difference in the order between the denominator and the numerator are at ∞. The zeros at $s = -\infty$ in the s-plane are mapped to $z = e^{-\infty}e^{-j(\pi/T)T} = -1$.
4. If a unit delay in the discrete domain transfer function is desired, then one zero of $H(s)$ at $s = -\infty$ is not mapped to the z-plane. This implies that the numerator of $H(z)$ will have one order less as compared to the denominator of $H(z)$.
5. The gain of $H(z)$ is selected to match the gain of $H(s)$ at a critical point like

$$H(s)\Big|_{s=0} = H(z)\Big|_{z=1}$$

Problem 9.1

Consider the following continuous domain transfer function

$$H(s) = \frac{a}{s+b}$$

Convert H(s) into the discrete domain.

Solution

Applying the rules 1 and 3 above, one obtains

$$H(z) = K_z\frac{z+1}{z-e^{-bT}}$$

Applying rule 5, one obtains

$$H(s)\Big|_{s=0} = \frac{a}{b} = H(z)\Big|_{z=1} = K_z\frac{2}{1-e^{-bT}}$$

From the above equation,

$$K_z = \frac{a}{b}\left(\frac{1-e^{-bT}}{2}\right)$$

Now, substituting for K_z in $H(z)$, one obtains

$$H(z) = \frac{a}{b}\left(\frac{1-e^{-bT}}{2}\right)\left(\frac{z+1}{z-e^{-bT}}\right)$$

Problem 9.2

Consider a system whose continuous domain transfer function is given by

$$H(s) = \frac{10(s+1)}{(s+2)(s+3)}$$

Convert H(s) into the discrete domain.

Solution

Applying rules 1, 2 and 3, one obtains

$$H(z) = K_z \frac{(z-e^{-T})(z+1)}{(z-e^{-2T})(z-e^{-3T})}$$

Applying rule 5, one obtains

$$H(s)\Big|_{s=0} = \frac{10}{6} = H(z)\Big|_{z=1} = K_z \frac{(1-e^{-T})2}{(1-e^{-2T})(1-e^{-3T})}$$

From the above equation,

$$K_z = \frac{5}{6}\frac{(1-e^{-2T})(1-e^{-3T})}{(1-e^{-T})}$$

Substituting for K_z in $H(z)$, one obtains

$$H(z) = \frac{5}{6}\frac{(1-e^{-2T})(1-e^{-3T})}{(1-e^{-T})}\frac{(z-e^{-T})(z+1)}{(z-e^{-2T})(z-e^{-3T})}$$

Numerical Integration Methods

Integration is a central idea in physical systems. The state equation models and the transfer function models all basically indicate the integration concept. The state variables are obtained by integration and the differential equations can be solved and simulated by using only the integrator block. This implies that any physical system may be modeled with just integrator blocks. Therefore, one needs to convert the continuous integrator model into the discrete domain. By applying the conversion to all the integrators in the system model, the discrete domain model is obtained. Thus, $1/s$ which is the transfer function of the integrator in the s-domain should be equivalently represented in the z-domain. Then the inverse of the discrete domain integrator is used to replace the "s" variable in the s-domain transfer function to obtain the z-domain transfer.

Consider a continuous signal $x(t)$ as shown in Figure 9.30. Passing this signal through an integrator means the area under the signal $x(t)$ is the output of the integrator.

Figure 9.31 gives the continuous signal $x(t)$ and the concept for finding the integration value of $x(t)$ by finding the area under the curve in the discrete domain.

Let $y(n)$ be the variable associated with the integrator output and $x(n)$ be the sampled version of the continuous signal $x(t)$ that is applied to the integrator input. Then at any instant k, $y(kT)$ or $y(k)$ contains the area under the curve. Note that $y(k-1)$ is the area under the curve till time instant $(k-1)T$. Therefore, the area under the curve may be written as

Area under curve till kT = Area under curve till $(k-1)T$ + Area under curve between time kT and $(k-1)T$

This can be written as

$$y(k) = y(k-1) + (\text{Area update}) \tag{9.11}$$

Equation (9.11) is the central theme that is used in all integration algorithms in the digital domain. This is illustrated in the block schematic of Figure 9.32. The block schematic of the discrete integrator shown in

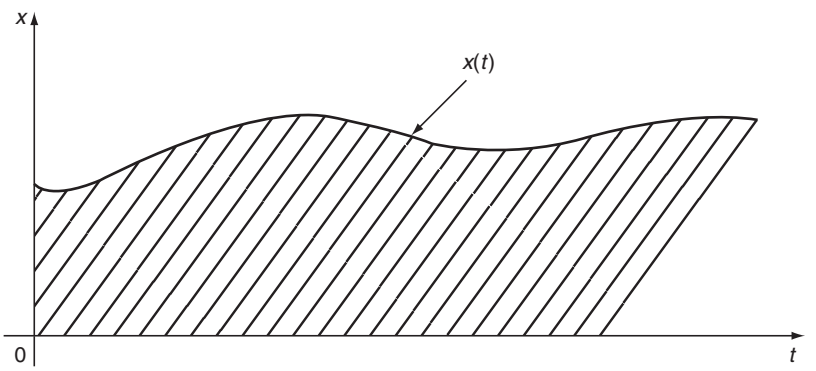


Figure 9.30 | A continuous signal $x(t)$ and the shaded region is the area under the curve.

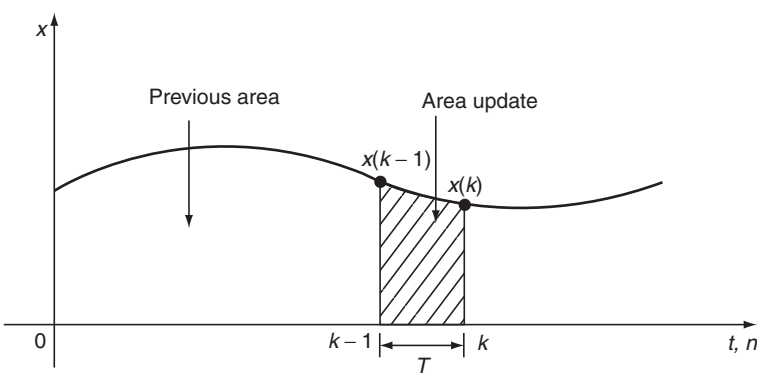


Figure 9.31 | Finding the area under the curve in the discrete domain.

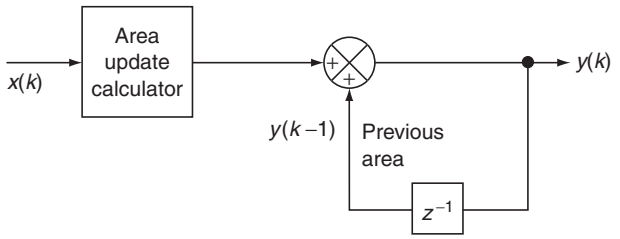


Figure 9.32 | Generic discrete integration block schematic.

Figure 9.32 is a very generic algorithm. The various integration algorithms in fact differ in the area update calculation method. Based on the generic block schematic of Figure 9.32, three algorithms, namely, Euler's forward rectangular rule, Euler's backward rectangular rule and the Tustin's rule, will be discussed in this section.

Euler's Forward Rectangular Rule

In this method, the update area is calculated by approximating the area of a rectangle of width T and height equal to the signal value at the previous instant of time, that is, $x(k-1)$. The height of the update rectangle is taken as $x(k-1)$ which is brought forward to the instant kT, hence the name forward rectangular rule. The update area is given by $x(k-1) \times T$. The method is illustrated in Figure 9.33.

Referring to the generic block schematic of Figure 9.32 and the Euler's forward rectangular illustration shown in Figure 9.33, one obtains

$$y(k) = y(k-1) + x(k-1)T$$

This can be expressed in the z-domain as

$$Y(z) = z^{-1}Y(z) + z^{-1}TX(z)$$

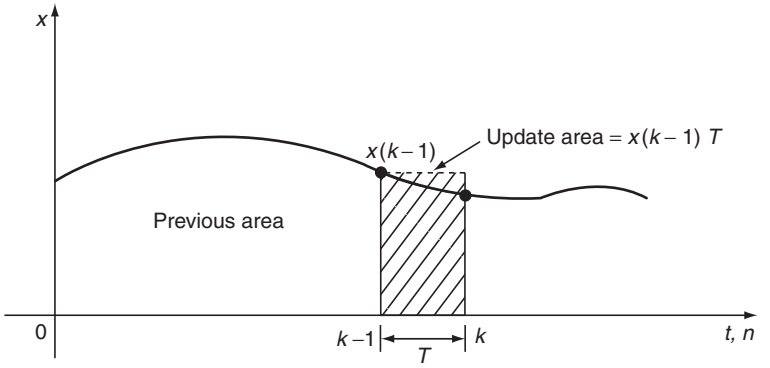


Figure 9.33 | Illustration of Euler's forward rectangular rule.

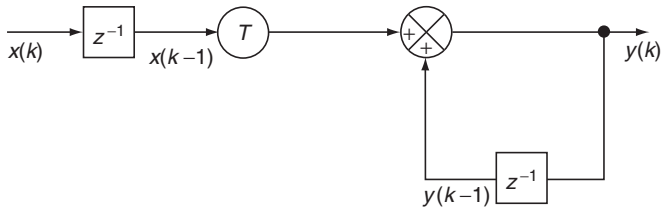


Figure 9.34 | Block schematic of Euler's forward rectangular rule.

From the above equation, the transfer function of the Euler's forward rectangular integrator is given as

$$H(z)=\frac{Y(z)}{X(z)}=\frac{T}{z-1} \tag{9.12}$$

The block schematic of the Euler's forward rectangular rule is shown in Figure 9.34. It is an IIR filter with $b_0 = 0$, $b_1 = T$ and $a_1 = 1$.

Referring to Figure 9.17, it was shown that $s = 0$ in the s-domain maps to $z = 1$ in the z-domain. In the s-domain the integrator transfer function is $1/s$ which implies a pole at $s = 0$ in the s-domain. Note that Eq. (9.12) has a pole at $z = 1$ in the z-domain which is expected as per the s-domain to z-domain mapping. It is important to note that all forms of integrator in the discrete domain will have the pole at $z = 1$. It is only the scalar gain and the zero location that will vary for different algorithms.

For the Euler's forward rectangular rule, the s-domain and z-domain integration equivalence is

$$\frac{1}{s}\leftrightarrow\frac{T}{z-1}\quad\text{or}\quad s\leftrightarrow\frac{z-1}{T} \tag{9.13}$$

From the above equivalence, the "s" variable is replace by $(z-1)/T$ in the continuous domain transfer function $H(s)$ to obtain the discrete domain transfer function $H(z)$. Therefore,

$$H(z)=H(s)\Big|_{s=\frac{z-1}{T}} \tag{9.14}$$

Equation (9.14) is read as: $H(z)$ equals $H(s)$ wherein the variable "s" is replaced by $(z-1)/T$ in the continuous domain transfer function.

Problem 9.3

Consider a continuous domain transfer function H(s) given as

$$H(s)=\frac{a}{s+b}$$

Convert H(s) into the discrete domain.

Solution

Applying Eq. (9.14), one obtains

$$H(z)=\frac{a}{\left(\frac{z-1}{T}\right)+b}=\frac{aT}{z-1+bT}$$

Euler's Backward Rectangular Rule

In this method, the update area is calculated by approximating the area of a rectangle of width T and height equal to the signal value at the present instant of time, that is, $x(k)$. The height of the update rectangle is taken as $x(k)$ which is taken back to the instant $(k-1)T$, hence the name backward rectangular rule. The update area is given by $x(k)T$. The method is illustrated in Figure 9.35.

Referring to the generic block schematic of Figure 9.32 and the Euler's backward rectangular illustration shown in Figure 9.35, one obtains

$$y(k) = y(k-1) + x(k)T$$

This can be expressed in the z-domain as

$$Y(z) = z^{-1}Y(z) + TX(z)$$

From the above equation, the transfer function of the Euler's backward rectangular integrator is given as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{Tz}{z-1} \tag{9.15}$$

The block schematic of the Euler's backward rectangular rule is shown in Figure 9.36. It is an IIR filter with $b_0 = T$, $b_1 = 0$ and $a_1 = 1$.

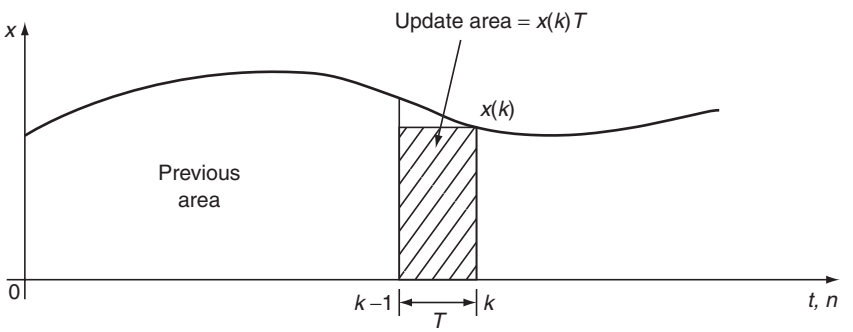


Figure 9.35 | Illustration of Euler's backward rectangular rule.

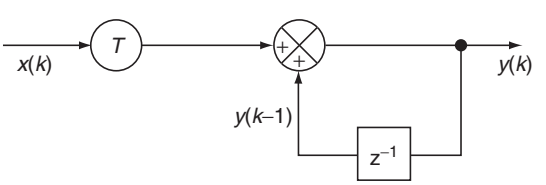


Figure 9.36 | Block schematic of Euler's backward rectangular rule.

Here also, note that Eq. (9.15) has a pole at $z = 1$ in the z-domain which is expected as per the s-domain to z-domain mapping. For the Euler's backward rectangular rule, the s-domain and z-domain integration equivalence is

$$\frac{1}{s} \leftrightarrow \frac{Tz}{z-1} \quad \text{or} \quad s \leftrightarrow \frac{z-1}{Tz} \tag{9.16}$$

From the above equivalence, the "s" variable is replace by $(z-1)/Tz$ in the continuous domain transfer function $H(s)$ to obtain the discrete domain transfer function $H(z)$. Therefore,

$$H(z) = H(s)\Big|_{s=\frac{z-1}{Tz}} \tag{9.17}$$

Problem 9.4

Consider the continuous domain transfer H(s) given as

$$H(s) = \frac{a}{s+b}$$

Convert H(s) into the discrete domain.

Solution

Applying Eq. (9.17), one obtains

$$H(z) = \frac{a}{\left(\frac{z-1}{Tz}\right)+b} = \frac{aTz}{z-1+bTz} = \left(\frac{aT}{1+bT}\right)\left(\frac{z}{z-\frac{1}{1+bT}}\right)$$

Tustin's Trapezoidal Rule

In this method, the update area is calculated by approximating the area of a trapezium as illustrated in Figure 9.37. Referring again to the generic block schematic of Figure 9.32 and the Tustin's trapezoidal rule illustrated in Figure 9.37, one obtains

$$y(k) = y(k-1) + \left(\frac{x(k-1)+x(k)}{2}\right)T$$

This can be expressed in the z-domain as

$$Y(z) = z^{-1}Y(z) + z^{-1}\frac{T}{2}X(z) + \frac{T}{2}X(z)$$

From the above equation, the transfer function of the Tustin's trapezoidal integrator is given as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{T}{2}\left(\frac{z+1}{z-1}\right) \tag{9.18}$$

Equation (9.18) is also called the Tustin's form or the trapezoidal integrator or bilinear transformation in the literature.

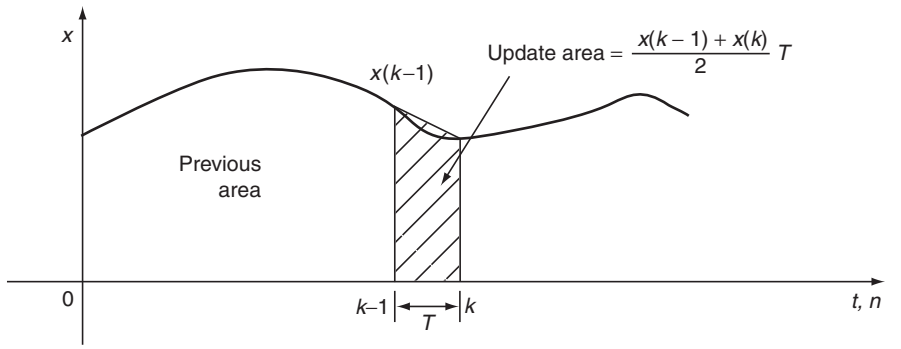


Figure 9.37 | Illustration of Tustin's trapezoidal rule.

The block schematic of the Tustin's trapezoidal rule is shown in Figure 9.38. It is an IIR filter with $b_0 = T/2$, $b_1 = T/2$ and $a_1 = 1$.

Here again note that Eq. (9.18) has a pole at $z = 1$ in the z-domain which is expected as per the s-domain to z-domain mapping. For the Tustin's trapezoidal rule, the s-domain and z-domain integration equivalence is

$$\frac{1}{s} \leftrightarrow \frac{T}{2}\frac{(z+1)}{(z-1)} \quad \text{or} \quad s \leftrightarrow \frac{2}{T}\frac{(z-1)}{(z+1)} \tag{9.19}$$

From the above equivalence, the "s" variable is replaced by $2(z-1)/[T(z+1)]$ in the continuous domain transfer function $H(s)$ to obtain the discrete domain transfer function $H(z)$.

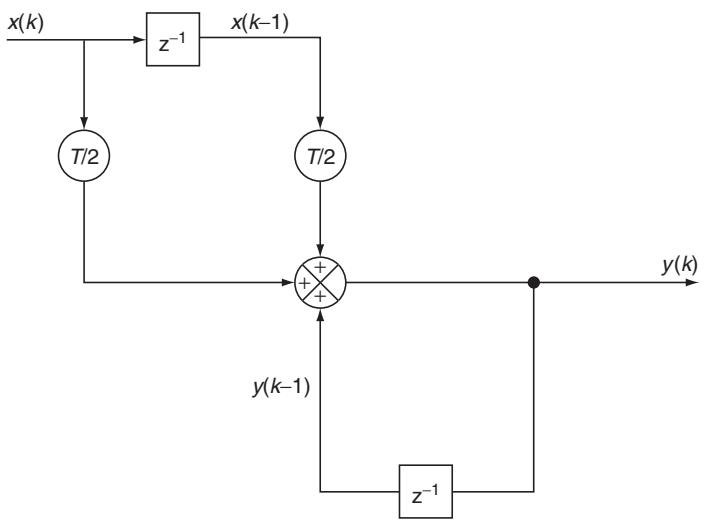


Figure 9.38 | Block schematic of Tustin's trapezoidal rule.

Therefore,

$$H(z) = H(s)\Big|_{s=\frac{2}{T}\frac{(z-1)}{(z+1)}} \tag{9.20}$$

Problem 9.5

Consider the continuous domain transfer H(s) given as

$$H(s) = \frac{a}{s+b}$$

Convert H(s) into the discrete domain.

Solution

Applying Eq. (9.20), one obtains

$$H(z) = \frac{a}{\frac{2}{T}\left(\frac{z-1}{z+1}\right)+b} = \left(\frac{aT}{2+bT}\right)\frac{z+1}{z-\left(\frac{2-bT}{2+bT}\right)}$$

Hold Equivalence

This is one of the most popular methods to convert continuous domain transfer function to the discrete domain equivalent as it is very similar to the practical scheme. Consider a system whose continuous domain transfer function is $H(s)$. It receives as input a continuous signal $x(t)$ and delivers as its output a continuous signal $y(t)$ as shown in Figure 9.39(a). The same system, when represented in the discrete domain as a transfer function $H(z)$, should receive a discrete signal $x(k)$ which is the sampled version of $x(t)$ and deliver as its output a discrete signal $y(k)$ which should be the sampled version of $y(t)$ as indicated in Figure 9.39(b). Then one can say that $H(z)$ is the discrete equivalent of $H(s)$. To convert the continuous system to a discrete system, one must include a DAC at the input of $H(s)$ and an ADC at the output of $H(s)$ as shown in Figure 9.39. This way the digital input $x(k)$ is converted to analog signal which is applied as input to the continuous system. The output of the continuous system is a continuous domain signal which is discretized by passing it through an ADC to deliver the discrete output signal $y(k)$.

The DAC in Figure 9.39(b) is called the hold device and the ADC is called the sampler. Therefore, the discrete equivalent model of the system can be re-drawn as shown in Figure 9.40. The hold, the continuous system and the sampler together form the discrete equivalence of the continuous system. This is called the *hold equivalence*. The DAC output is held constant till the next sampled value. This type of hold is called the zero-order hold. Hence this type of equivalence is also called the *zero-order hold equivalence*.

The hold device accepts a sample $x(k)$ at $t = kT$ and holds its output constant until the next sample is received at $t = (k+1)T$. The piece-wise constant output of the hold device is the signal that is applied to the system.

Consider a unit impulse being applied to the input of the hold device at instant kT. The output of the hold device is a pulse of amplitude 1 and width T as shown in Figure 9.41. The output of the hold device can be considered to be the difference between a step at kT and a delayed step at $(k+1)T$. The Laplace transform of a unit step is given as $1/s$. *Laplace transform of the step response of the system is H(s)/s and the Laplace transform of the delayed step response of the system is* $[H(s)/s]e^{-st}$ *as a unit delay* z^{-1} *in the discrete domain is* e^{-sT} *in the s-domain from Eq. (9.9).*

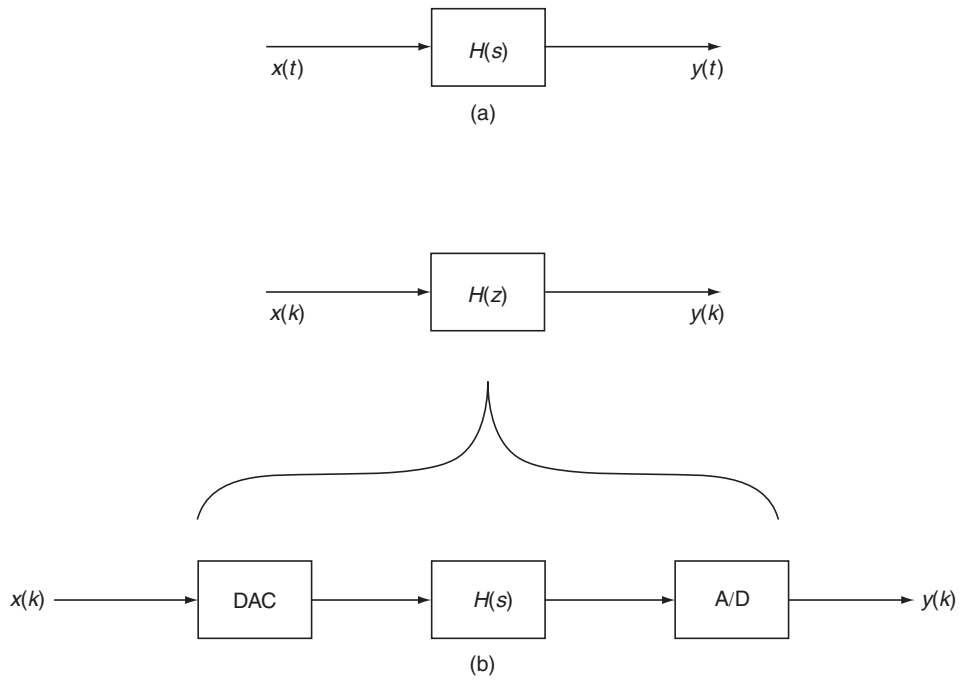


Figure 9.39 (a) Continuous domain transfer function; (b) discrete domain transfer function equivalence.

As the system is linear, superposing both these responses, the response of the system to a unit impulse at the input of the hold is given by

$$Y(s) = \frac{H(s)}{s} - \frac{H(s)}{s} e^{-sT} = (1 - e^{-sT}) \frac{H(s)}{s} \tag{9.21}$$

The response of the system $y(t)$ is the Laplace inverse of $Y(s)$. The $y(t)$ is passed through the sampler to be delivered as $y(k)$ which is the discrete impulse response of the "hold equivalence" system shown in Figure 9.40.

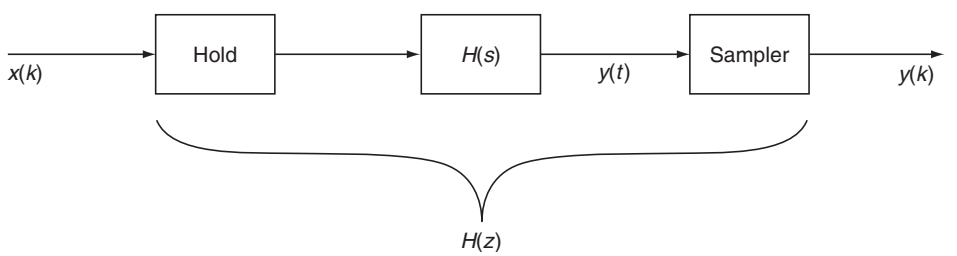


Figure 9.40 Discrete transfer function by hold equivalence.

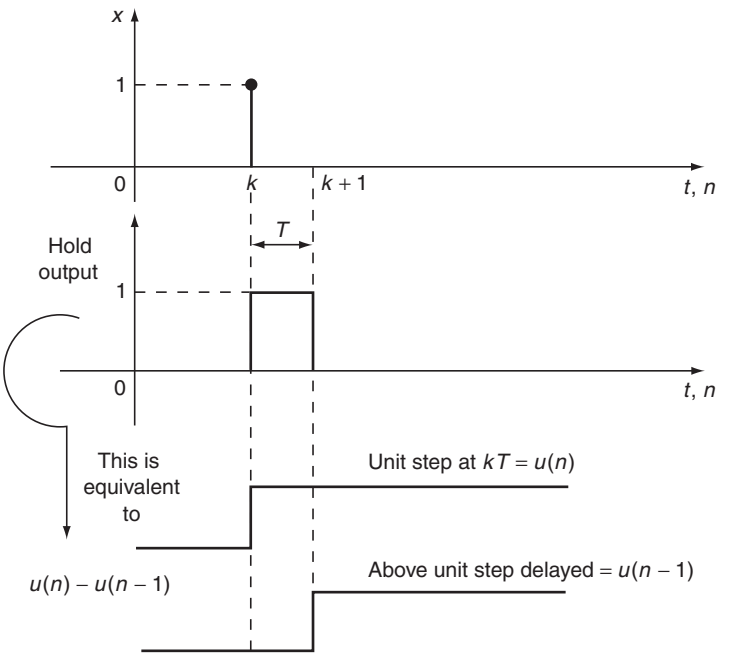


Figure 9.41 | Unit impulse response of the hold device.

Thus,

$$H(z) = \mathbb{Z}\{Y(s)\} \tag{9.22}$$

where $\mathbb{Z}$ is the z-transform operator. Therefore, from Eqs. (9.21) and (9.22), one obtains

$$H(z) = \mathbb{Z}\left\{(1-e^{-sT})\frac{H(s)}{s}\right\} = (1-z^{-1})\cdot\mathbb{Z}\left\{\frac{H(s)}{s}\right\} \tag{9.23}$$

Equation (9.23) gives the zero-order hold discrete transfer function equivalence of the continuous system $H(s)$. This is also simply referred to as ZOH equivalence. The $\mathbb{Z}\{H(s)/s\}$ portion of $H(z)$ is found by referring to the z-transform tables given in Appendix IV. However, the ZOH equivalence is difficult to convert manually. It is better done using a computer. MATLAB™ and SciLAB provide functions to perform the continuous to discrete domain conversions.

Problem 9.6

Consider the continuous transfer function

$$H(s) = \frac{a}{s+b}$$

Convert H(s) into the discrete domain.

Solution

Given

$$H(s) = \frac{a}{s+b}$$

Therefore

$$\frac{H(s)}{s} = \frac{a}{s(s+b)} = \frac{a}{b}\left(\frac{1}{s} - \frac{1}{s+b}\right)$$

Referring to the z-transform tables in Appendix IV,

$$\mathbb{Z}\left\{\frac{H(s)}{s}\right\} = \frac{a}{b}\left(\frac{z}{z-1} - \frac{z}{z-e^{-bT}}\right) = \frac{az(1-e^{-bT})}{b(z-1)(z-e^{-bT})}$$

Applying the ZOH equivalence of Eq. (9.23), one obtains

$$H(z) = \frac{a}{b}\left(\frac{1-e^{-bT}}{z-e^{-bT}}\right)$$

Discretizing the State Equation

As discussed in Chapter 8, a continuous system can be represented by a set of first-order linear differential equations. This set can be expressed in a standard matrix form called the state equation as given in Eq. (9.24) and the output equation as given in Eq. (9.25):

$$\dot{x} = A \cdot x + B \cdot u \tag{9.24}$$

$$y = C \cdot x + D \cdot u \tag{9.25}$$

where x is the vector of state variables, u is the vector of input variables and y is the vector of output variables. A is called the state matrix, B the input weighting matrix, C the output matrix and D the input–output direct feed-forward matrix.

The continuous state equation representation as given in Eqs. (9.24) and (9.25) needs to be converted to the discrete state equation representation which is of the form

$$x(k+1) = A_d x(k) + B_d u(k) \tag{9.26}$$

$$y(k) = C_d x(k) + D_d u(k) \tag{9.27}$$

Equation (9.26) is the discrete state equation representation and Eq. (9.27) is the discrete output equation. The continuous domain state equation is a set of linear differential equations whereas the discrete domain state equation is a set of difference equations. Equation (9.24) gives the slope or the gradient of the state variables based on which the future evolution of the state variables can be predicted whereas the discrete state equation [Eq. (9.26)] gives the prediction of the next value of the state variables. The output equations [Eqs. (9.25) and (9.27)] are identical in both the cases as the output is the linear combination of the state and input variables which are the same whether in continuous domain or in discrete domain. Therefore,

$$C_d = C \text{ and } D_d = D \tag{9.28}$$

It is now required to obtain a conversion relation for A_d and B_d in terms of the continuous domain matrices and sampling time T. Consider the set of linear differential equation [Eq. (9.24)]. The solution of this gives the state vector. The solution of Eq. (9.24) has two parts: (a) the complementary function which is the general solution and (b) the particular integral which is the particular solution to a specific input.

The complementary function solution of Eq. (9.24) is

$$e^{A(t-t_o)}x(t_o)$$

where t_o is the starting time and $x(t_o)$ is the initial state value and the particular integral or the particular solution is given by

$$\int_{t_o}^{t} e^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau$$

The total solution is the sum of the complementary function and the particular solution. Thus,

$$x(t) = e^{A(t-t_o)}x(t_o) + \int_{t_o}^{t} e^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau \tag{9.29}$$

Let $x(t)$ be passed through a sampler of sampling period T. Consider two samples of $x(t)$, one at kT which is the starting time representing t_o in Eq. (9.29) and the other at $t = (k+1)T = kT + T$, which is the value one sample period later. Thus, the sampled version of Eq. (9.29) becomes

$$x(kT+T) = e^{AT}x(kT) + \int_{kT}^{kT+T} e^{A(kT+T-\tau)}Bu(\tau)\mathrm{d}\tau \tag{9.30}$$

If a zero-order hold is used for sampling and holding the input signal $u(\tau)$, then

$$u(\tau) = u(kT), \qquad kT \le \tau < kT + T$$

Here $u(kT)$ will be a constant during the period kT to $(k+1)T$ due to the use of zero-order hold. Now change the integration variable in Eq. (9.30) from τ to t such that

$$t = kT + T - \tau$$

The discrete solution given in Eq. (9.30) becomes

$$x(kT+T) = e^{AT}x(kT) + \left(\int_{0}^{T} e^{At}\mathrm{d}tB\right)u(kT) \tag{9.31}$$

Comparing Eq. (9.31) with the discrete state equation standard form given in Eq. (9.26), one obtains

$$A_d = e^{AT} \tag{9.32}$$

$$B_d = \int_{0}^{T} e^{At}\mathrm{d}tB \tag{9.33}$$

and from Eq. (9.28), one has

$$C_d = C \tag{9.34}$$

$$D_d = D \tag{9.35}$$

Equations (9.32)–(9.35) give the relationship to convert the continuous state space model to the discrete state space model. These conversion equations are tedious to do by hand. They are best done with the help of computer using software tools like MATLAB™ or SciLAB.

9.7 Control System Basics

This section discusses a few common concepts and terminologies that will be useful in understanding the control concepts that will be used in the succeeding chapters of this book.

Performance Specifications

Specifications are essential to quantify the performance of the control system. These are called the performance specifications. Analysis and design are made with respect to these performance specifications. A control system may be qualified with the following three performance parameters:

1. Stability;
2. Steady-state accuracy;
3. Speed of response.

These are called the SSS performance parameters. The control system should be specified with respect to the above SSS performance parameters. The quantification of these parameters is called the performance specifications.

Stability

Consider a system which is applied by a step signal at the input. Due to the applied step disturbance at the input, three possibilities can happen to the output response of the system. (a) The step input will disturb the output response in such way that the output signal will grow and go unbounded as shown in Figure 9.42(a). Such a system is called unstable. (b) The step input will disturb the output response and output will oscillate without decaying as shown in Figure 9.42(b). Such a system is called marginally stable. (c) The step input will disturb the output response; however, the output disturbance will asymptotically decay to zero as shown in Figure 9.42(c). Such a system is called stable or asymptotically stable.

Figure 9.42 gives the time domain view point about the stability of the system. The pole-zero domain view point is given in Figure 9.43. Consider the pole-zero domain or the s-domain shown in Figure 9.43. In Figure 9.43(a), a pole of a system is on the real positive axis at "a". This has a transfer function $1/(s-a)$. The time response of this transfer function is e^{at}. As $a > 0$ and $t > 0$, the response of the system continuously increases with time, that is, the response goes unbounded. Such a system is called an unstable system. All the

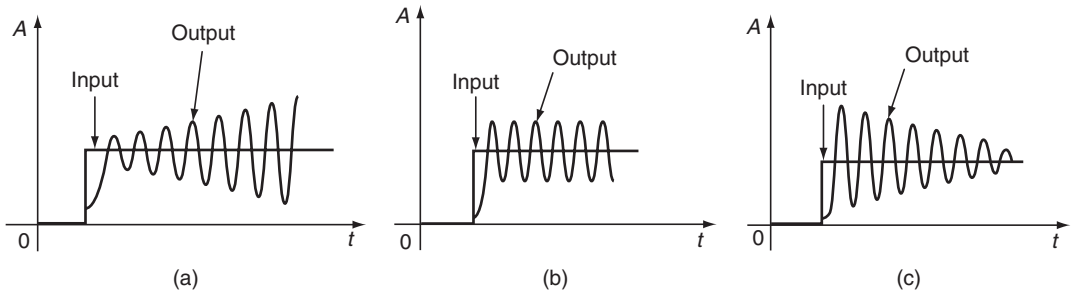


Figure 9.42 Time domain view point of stability: (a) Unstable system; (b) marginally stable system; (c) asymptotically stable system.

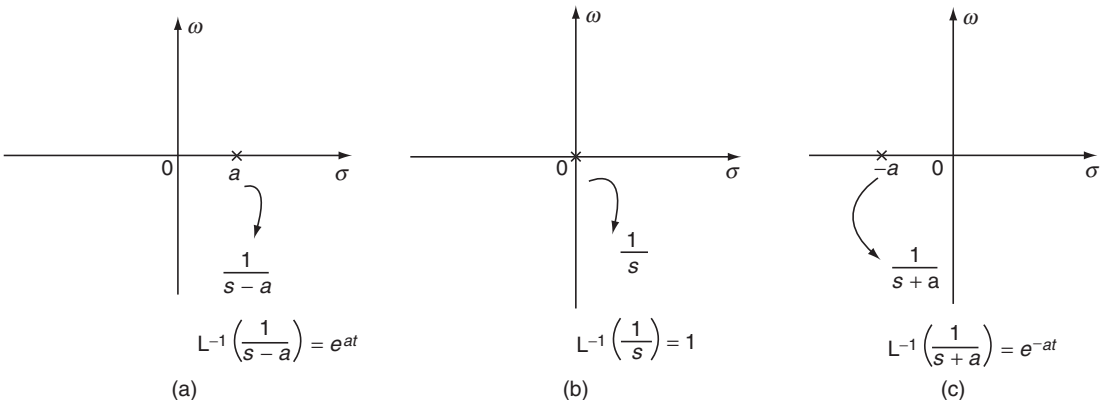


Figure 9.43 Pole-zero domain view point of stability: (a) Unstable system; (b) marginally stable system; (c) stable system.

poles that are on the right half of the *s*-plane will give a system response that goes unbounded. Therefore, all the systems that *have pole or poles located on the right half* of the *s*-plane are considered as unstable systems. By Eq. (9.9), this implies that all discrete systems having poles located outside the unit circle in the *z*-plane are unstable systems.

In Figure 9.43(b), the pole of the system is located at $s = 0$. This is an integrator. Here the response is not a function of time. Here it neither increases with time nor decreases with time. Even in the integrator it is evident that on giving an input pulse and removing the pulse, the integrator output will neither decay nor increase with time. Such a system is called marginally stable. All systems that have poles *only on the ω-axis* of the *s*-plane will have a response that will neither increase nor decrease with time. They may oscillate at the frequency where the pole is located on the ω-axis of the *s*-plane, but the amplitude of the oscillations will neither increase nor decrease. By Eq. (9.9), this implies that all discrete systems having poles located on the unit circle in the *z*-plane are marginally stable systems.

In Figure 9.43(c), the pole of the system is located at $s = -a$. Here the response is given as e^{-at}. This implies that the system response decays with time which implies that if any disturbance occurs, the disturbance in the response will exponentially or asymptotically decay to zero with time. Such a system is a stable system. All systems that have their poles located *in the left half of the s-plane* only will possess a decaying mechanism and hence are called the stable systems. By Eq. (9.9), this implies that all discrete systems having poles located within the unit circle in the *z*-plane are stable systems.

Steady-State Accuracy

The system response when steady state has been reached is an important performance criterion. The system will in fact spend a major portion of its operating time in the steady state; therefore the error that is present in this state will give a measure of the quality of the system. Consider the closed-loop system shown in Figure 9.44. The system has a continuous transfer function in the *s*-domain, $G(s)$. The output is fed back with unity gain. The reference is compared with the fed back signal, that is, the output to obtain the error signal which is applied to the input of the system.

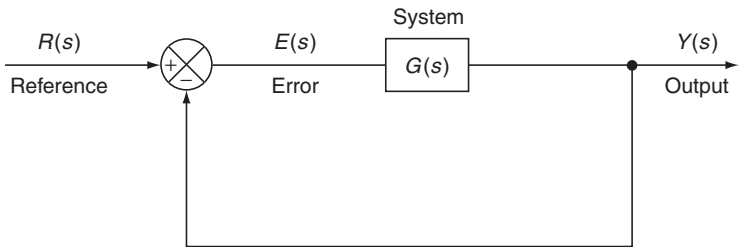


Figure 9.44 | Closed-loop system for steady-state error analysis.

Let $G(s)$ be expressed in a generalized manner in the pole-zero form as

$$G(s) = \frac{K(1+T_a s)(1+T_b s)\ldots}{s^m(1+T_1 s)(1+T_2 s)\ldots} \tag{9.36}$$

The error in the s-domain $E(s)$ is given by

$$E(s) = \frac{R(s)}{1+G(s)} \tag{9.37}$$

The number "m" of pure integrations determines the *type* of the system. If m is zero, the system is a type 0 system; if m is 1 then the system is a type 1 system and so forth.

The standard test inputs for a system are steps, ramps and parabolic-shaped signals. Steady-state errors due to these standard test inputs are determined in terms of constants called the *error constants*. These error constants give a measure of the quality of the system's steady-state performance. For any system with a generalized transfer function $G(s)$ as given in Eq. (9.36), the error constants are defined as follows:

Positional error constant

$$K_{\mathrm{p}} \triangleq \lim_{s\to 0} G(s) \tag{9.38}$$

For the discrete domain system it is

$$K_{\mathrm{p}} \triangleq \lim_{z\to 1} G(z)$$

Velocity error constant

$$K_{\mathrm{v}} \triangleq \lim_{s\to 0}[sG(s)] \tag{9.39}$$

For the discrete domain system it is

$$K_{\mathrm{v}} \triangleq \lim_{z\to 1}[(1-z^{-1})G(z)]$$

Acceleration error constant

$$K_{\mathrm{a}} \triangleq \lim_{s\to 0}[s^2 G(s)] \tag{9.40}$$

For the discrete domain system it is

$$K_{\mathrm{a}} \triangleq \lim_{z\to 1}[(1-z^{-1})^2 G(z)]$$

Table 9.3 | Steady-state errors for various system types

	Unit step, $R(s) = 1/s$	**Ramp, $R(s) = 1/s^2$**	**Parabolic, $R(s) = 1/s^3$**
System type	$e_{ss} = \dfrac{1}{1+K_p}$	$e_{ss} = \dfrac{1}{K_v}$	$e_{ss} = \dfrac{1}{K_a}$
0	$1/(1+K)$	∞	∞
1	0	$1/K$	∞
2	0	0	$1/K$
$m \geq 3$	0	0	0

By using the final value theorem on $E(s)$ as given in Eq. (9.37), the steady-state error e_{ss} is given by

$$e_{ss} = \lim_{t\to\infty} e(t) = \lim_{s\to 0}[sE(s)] = \lim_{s\to 0}\left[\frac{sR(s)}{1+G(s)}\right] \tag{9.41}$$

For the discrete domain system

$$e_{ss} = \lim_{t\to\infty} e(t) = \lim_{z\to 1}[(1-z^{-1})E(z)] = \lim_{z\to 1}\left[\frac{(1-z^{-1})R(z)}{1+G(z)}\right]$$

Table 9.3 summarizes the steady-state errors for systems of various types and various standard inputs.

From Table 9.3 it may be seen that for a type 0 system, the steady-state error is finite for a unit step input. For a type 1 and above system, the steady-state error for a unit step input is zero whatever may be the value of the gain K. The steady-state error constants have similar meaning for discrete domain systems too.

Speed of Response

The speed of response of a system is the quickness with which it can reach the steady state on the occurrence of disturbances or input changes. This is generally quantified in terms of the rise time of the output signal. When a unit step is given to the input of the system, the output varies from the initial to the final settling value. The time taken for the output to change from 10% of the final steady-state value to 90% of the final steady-state value is a measure of the speed of response in the time domain. In the frequency domain, the bandwidth of the system gives a measure of the speed of response of the system. If the speed of response is high then the bandwidth in the frequency domain is high and vice versa.

Measures for Performance Parameters

Consider a system with reference input $r(t)$ and output response $y(t)$. The input is a unit step and the typical output signal response to a unit step input is shown in Figure 9.45. Many performance parameters are identified in the output response shown in Figure 9.45.

1. **Stability Measure:**

$$P_o = \frac{y_{max} - y_{final}}{y_{final}} \times 100$$

where P_o is the peak overshoot. This parameter gives a measure of the relative stability. If the value of P_o is low then the relative stability is high and vice versa.

2. **Steady-State Accuracy Measures:** e_B is called the error band when the system has settled down or reached its steady state. e_B gives a measure of the steady-state accuracy.
3. **Speed of Response Measures:** Figure 9.45 shows many parameters that are measures of speed of response. The time taken by the output response to stay within the error band e_B is called the settling time (t_s). The time taken by the output response to rise from 10% to 90% of the final value is called the rise time (t_r) of the output response. The time taken by the output response to reach 50% of its final value is called the delay time (t_d). Note that t_s, t_r and t_d are measures of speed of response. Also, t_r and t_d give the speed of the initial dynamics. However, if one is interested in the dynamics upto steady state then t_s is a better measure.

Consider a system having the *s*-domain closed-loop transfer function given by $G(s)$ or a *z*-domain closed-loop transfer function given by $G(z)$. The transfer function in the ω-domain is denoted as $G(j\omega)$. The closed-loop frequency domain viewpoint of the output response is shown in Figure 9.46. Here again, one can visualize the SSS performance measures using frequency domain parameters.

1. **Stability Measure:** The maximum value of the magnitude of the closed-loop transfer function in the ω-domain M_p is a measure of the relative stability of the system. Higher the value of M_p, lower is the relative stability of the system.
2. **Steady-State Accuracy Measure**: The magnitude value of the ω-domain transfer function $G(j\omega)$ at $\omega = 0$ gives the closed-loop gain at steady state. If steady-state error needs to be zero, the magnitude of $G(j\omega)$ at $\omega = 0$ should be unity. Only then will the output be same as the input. The deviation of $G(j\omega)$ from unity at $\omega = 0$ gives a measure of the steady-state accuracy of the closed-loop response.

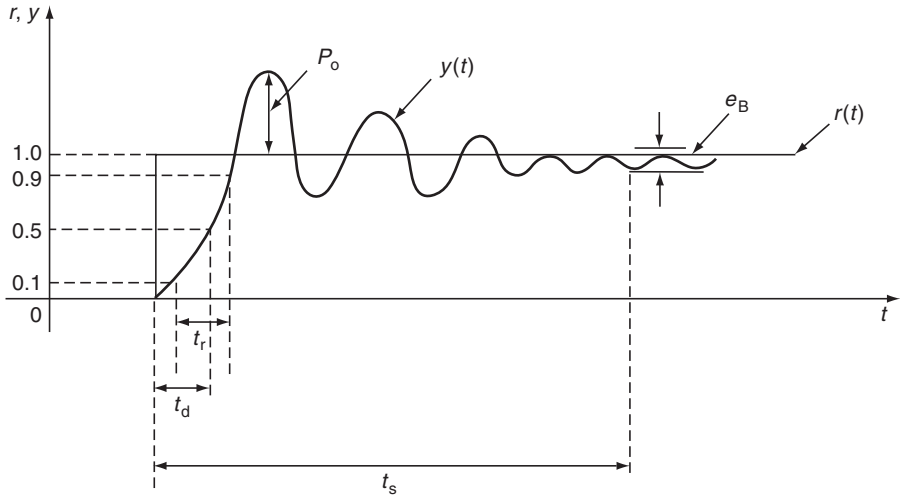


Figure 9.45 | Time domain performance measures for unit step standard input.

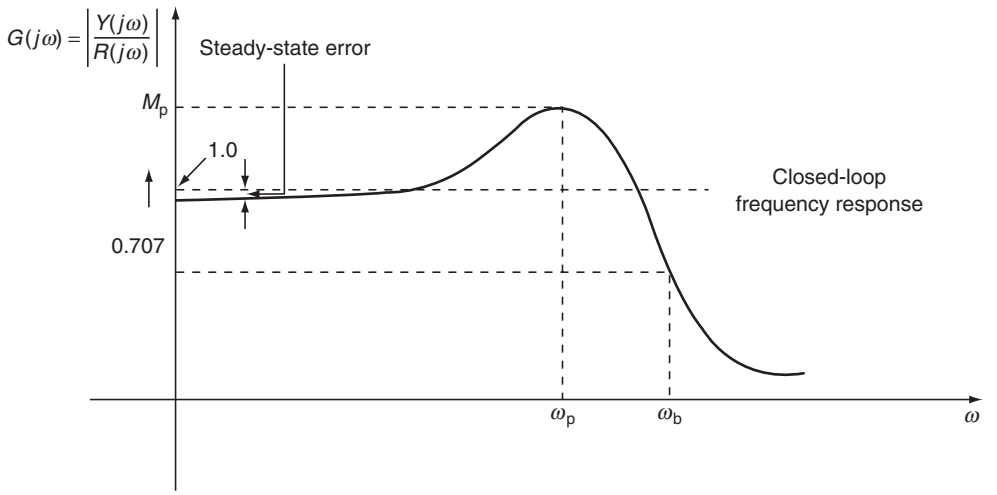


Figure 9.46 | Frequency domain visualization of performance measures of a closed-loop system.

3. **Speed of Response Measure:** ω_b gives the bandwidth of the system. At $\omega = \omega_b$, the magnitude of $G(j\omega)$ has a value of 0.707. This gives upper bound on the high-frequency components in the time domain signal and hence a measure of the rise time. Therefore ω_b can be considered to be a measure of the speed of response of the system.

9.8 Control Principles

This section will give a brief discussion on the basic control principles that will enable one to understand the concepts to be discussed in the forthcoming chapters. Consider the closed-loop system shown in Figure 9.47.

In Figure 9.47, the plant is controlled by a controller that is represented as a gain $K(\omega)$ which is a function of frequency ω. The core function of this controller K is to ensure that the error e is *always* zero. By *always*, it is meant that both during steady-state conditions and during transient conditions. The controller K generates a signal V_c such that error e is always kept at 0. Now the question to be answered is: At what value of K is e always 0?

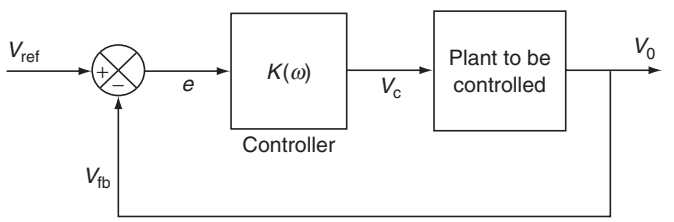


Figure 9.47 | Block schematic of a closed-loop system.

It is evident from Figure 9.47 that

$$e = \frac{V_c}{K} \tag{9.42}$$

From Eq. (9.42) it is evident that e is always 0 when

1. V_c is always zero or
2. K is always infinity (∞).

Condition (1) above implies open-loop operation. As V_c is always zero there is no control input to the plant. As a consequence, one loses control of the plant and the plant operates in open-loop condition. Condition (2) is the acceptable condition. If K is made ∞, then V_c is adjusted in such a manner that error e is always 0 under all conditions of disturbances and plant parameter variations. For any given system just choose K as infinite; then the error is always zero. K can be made infinite practically by using an op-amp as a comparator. This would be an ideal situation. However, this is not possible due to the following two main reasons:

1. The value of V_c is limited.
2. The noise in the system also gets amplified by K.

As all controllers and controlled systems need electronics, they need to be powered by power supplies that are either 3.3 V or 5 V or ±15 V. Therefore V_c cannot increase more than the power supply voltage. Hence, even if K is made infinite, V_c can take only finite values and therefore, if a value of the product of e and K is greater than the power supply voltage, then V_c can at most clamp to power supply voltage. This implies that V_c will not change for any change in e and therefore there will be loss of control or in other words, the system will be operating in open loop. The error band within which there is a change in V_c for a change in e is called the control band. Close-loop operation exists only if the operation is within the control band. Figure 9.48 shows the control band. From Eq. (9.42), it follows that

$$K = \frac{V_c}{e} \tag{9.43}$$

The plot of error e versus control voltage V_c is shown in Figure 9.48. The slope of the e–V_c curve is the controller gain K at that point. Consider that the gain K is constant over the error range. Then e and V_c will have a linear relationship with slope equal to K. Referring to Figure 9.48, consider that V_c can swing only between −15 V and +15 V. When the gain is K_1, the error e could swing between –1 and 1. If the error goes beyond ±1, then V_c will saturate to ±15. This range of error values where V_c does not saturate is called the control band. At higher gains, the control band decreases. Evidently, at $K = \infty$, the control band is 0. This means that if there is even a finite error due to disturbance or parameter changes, the control signal V_c will saturate. If V_c saturates, then the loop is broken and behaves like open loop due to loss of control on plant. The error would swing +ve and V_c goes to +ve saturation. This will lead to –ve error due to plant output swinging to a larger positive value. When the error goes negative, V_c will swing to –ve saturation and so on. This phenomenon is called limit cycle oscillation. This occurs due to practical limitations on the value of V_c. If V_c is capable of taking on values from $-\infty$ to $+\infty$ then one can achieve ideally zero error condition always.

The second practical problem is that of noise. In any electronic circuit, every component or device will generate noise like thermal noise, shot noise and other noises associated with semiconductor components. Apart from these there will also be noise and disturbance that will be induced from neighboring equipments and also through the power supply. Therefore, in any practical circuit, noise is prevalent. This noise will also get multiplied by K and appear at V_c. If K is infinite then V_c primarily contains only noise. As noise is predominant at higher frequencies, K should be made infinite at lower frequencies and very low at higher frequencies so that

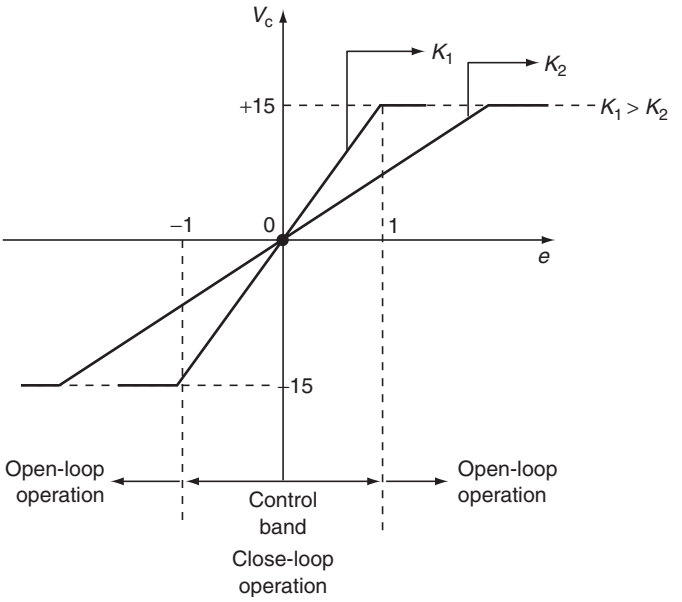


Figure 9.48 | Control band.

noise gets attenuated. Therefore, K should be a function of frequency ω. However, note that if K is made small at high frequencies then the bandwidth is limited which makes the speed of response lower. Thus the whole gamut of different control system techniques is basically for analyzing this gain $K(\omega)$ and to shape this frequency-dependent gain such that noise is attenuated but the performance specifications are met.

9.9 State Space Method

Traditionally, the classical control theory was employed for the design of the controllers of single-input single-output (SISO) systems. The classical control notions are generally in the frequency domain and the pole-zero domain. Relying on the transform methods, it is primarily applicable for linear time-invariant (LTI) systems. However, an exact description of the internal system dynamics is not needed for the classical design. Only the input–output behaviors of the system are of importance. A real system has disturbances and measurement noise that may not be possible to describe exactly by a mathematical model. Classical theory is natural for designing control systems that are robust to such disorders, yielding good closed-loop performance even though the controllers are not optimal with respect to a specific performance index. By robustness, it is meant that the system retains its performance and stability even in the face of disturbances and parameter variations in the system.

However, the classical control theory is difficult to apply in multiple-input and multiple-output (MIMO) systems like the case of a high performance induction motor vector-controlled drive system. Due to the interaction of the control loops in a multivariable system, each SISO transfer function can have acceptable properties in terms of step response and robustness, but the coordinated control motion of the system can fail to be acceptable. Therefore, MIMO design using classical techniques requires painstaking effort using

the approach of *closing one loop at a time*. This is a trial and error procedure that may require multiple iterations and it does not guarantee good results or even closed-loop stability.

On the other hand, by using state space control techniques, many of the limitations of the classical controls for multi-variable feedback control systems can be overcome. State space control design is fundamentally a time-domain technique. A state space model of the system to be controlled is required. The state space model can represent both MIMO as well as SISO systems. State space control techniques were first firmly established for linear systems and then later extended to non-linear systems. Topics like optimal and robust control are also time domain methods that essentially depend on the state space techniques.

Some remarks can be made on the features of the state space approach: (a) As long as the system is controllable, the feedback gains guarantee the stability of the closed-loop system. (b) This technique is easy to apply even for multiple-input plants. (c) The approach relies on the solution of the matrix design equations which are unsuited for hand calculations. As a consequence, computer-aided design is an essential feature. (d) All the controller gains of all the loops are determined at the same time by using the matrix design equations. Thus all the loops of a MIMO system are closed simultaneously which is in complete contrast to the one loop at a time procedure of classical controls design.

If one were to consider for example optimal controller like the linear quadratic regulator (LQR). Linear quadratic design using state feedback guarantees closed-loop stability. However, all the state components are seldom available for feedback purposes in a practical design problem. As a consequence, the non-available state components are estimated from the available states and the system model, using a reduced order observer. But unfortunately, the state estimates will depend on the accuracy with which the system parameters, that is, the elements of the A matrix are known. To overcome these problems, the output feedback method is also popular. In the output feedback approach, only the measurable or the available state components of the system are used for feedback control. Further, unlike the full-state feedback, the output feedback control can be used to design compensators with any desired dynamical structure, thereby regaining much of the intuition of classical controls design. These concepts are discussed in the succeeding two chapters.

CONCLUDING REMARKS

At the outset, this chapter may appear to be apparently disconnected with power electronics systems. However, as stated earlier in the introduction, this chapter is a cocktail of concepts and principles that are needed to better understand and design the controllers for the power electronics systems. For the design of controllers, first the physical power electronics system is modeled using the methods discussed in the previous chapter. The continuous domain model of the system is required to be shifted to the discrete domain in order to design the digital controllers. For this, the continuous to discrete domain conversion tools are required. Various methods have been discussed, but the zero-order hold or ZOH is by far the most popular.

To understand the discrete domain system representations, the z-transform, digital filters and sampling aspects are discussed. The mapping between the s-plane and the z-plane will give a visual comparative insight between the continuous and the discrete domain representations. Performance specifications and performance measures are introduced in the later part of the chapter to provide the framework for specifying the control-related measures. Finally, the next section gives few tutorial exercises that should be studied by simulating in a programming environment like SciLAB or MATLAB. These exercises will strengthen the discrete domain and control basics that are needed for the controller design which will be discussed in the next chapter.

TUTORIAL EXERCISES

The following tutorial exercises can be studied in either SciLAB/SciCos or MATLAB/Simulink simulation environment.

1. Construct a close-loop system with plant

$$\frac{(s+1)(s+2)}{(s+1.5)(s+2.5)(s+3.5)}$$

and a gain block K as the controller. Give a step input as the reference. Probe the error waveform and the control input to the plant.

- **a.** Vary K from 0.1 to 1000 and observe the waveforms.
- **b.** Keeping K at 0.1, introduce an integrator between the controller and the plant. Observe the waveforms.
- **c.** Introduce a limiter between the controller output and the plant. Repeat Steps (a) and (b).
- **d.** Introduce some high-frequency noise before the plant by including a random noise generator in the simulation schematic. Repeat Steps (a) and (b) above.
- **e.** Modify the plant by removing the zeroes. Now conduct Steps (a) and (b).

2. A system has a transfer function given by

$$H(z) = 1 + 5z^{-1} - 3z^{-2} + 2.5z^{-3} + 5z^{-8}$$

Determine and plot the output sequence $y(n)$ when the input is $\delta(n)$.

3. A system is described by the following input–output relation

$$y(n) = [x(n) + x(n-1) + x(n-2)]/3$$

Determine the system transfer function $H(z)$ for the system. Plot the poles and zeroes of $H(z)$ in the z-plane. What is the output if the input is $x(n) = 4 + \cos[0.25\pi(n-1)] - 3\cos(2\pi n/3)$?

4. Sample a sinusoidal waveform of 50 Hz using a 1 kHz sampler and plot the sampled sequence. Vary the sampling frequency from 100 Hz to 5 kHz and plot the frequency spectrum. What is effect of sampling frequency on the frequency spectrum?

5. The correct mapping between the s-plane and the z-plane is given by $z = e^{sT}$. Graphically give the s-plane to z–plane mapping along with the stability boundary for the following methods:

- **a.** Euler's forward rule
- **b.** Euler's backward rule
- **c.** Tustin's rule
- **d.** Zero-order hold

6. A digital filter has two poles at $z = 0$ and two zeros, one at $z = i$ and the other at $z = -i$. To this system apply a sine wave input of frequency 25 Hz which is sampled at 100 Hz. Build the above system in the simulation environment and observe the input and output on the scope. Also observe the frequency response of the above filter. Discuss and comment on the observations. What happens to the output when there is a slight deviation in the frequency of the sine wave about 25 Hz? Why?

7. A digital filter is designed such that its output is the average of the current input sample together with previous $(L-1)$ input samples. Write a function for the above L-point averaging FIR filter. The function has the following features:

- **a.** Input the value "L" for L-point averaging.
- **b.** Output the pole-zero map of the L-point average.
- **c.** Output the magnitude and phase frequency response.

8. Determine the range of controller gain K for stability of the unity feedback control system whose open-loop transfer function is

$$G(s) = \frac{K}{s(s+1)(s+2)}$$

9. Obtain the velocity error coefficients of the following systems through simulation:

a. $G(s) = \dfrac{10}{(s+1)(5s^2+2s+10)}$

b. $G(s) = \dfrac{3s+10}{5s^2+2s+10}$

10. Consider the unity feedback control system whose open-loop transfer function is

$$G(s) = \frac{100}{s(0.1s+1)}$$

Determine the steady-state error when the input is $r(t) = 1 + t + at^2$ where $a \geq 0$.

FILL IN THE BLANKS

1. Majority of the controllers exist in the _____ domain whereas the plant itself is in the _____ or the _____ domain.

2. A system exists always in the _____ domain.

3. The system dynamics and waveshape best viewed in the _____ domain.

4. Information like bandwidth, harmonic contents, spectrum, energy spectrum and power spectral density are better understood when viewed in the _____ domain.

5. The _____ domain provides information about damping, natural frequencies and dominant time constants for analysis and design.

6. To represent sampling of a continuous signal $x(t)$, one needs to multiply $x(t)$ with a train of _____.

7. There exists one-to-one correspondence between the signal representation in the _____ and the _____.

8. The co-efficients are _____ in both n- and z-domains.

9. Passing a signal through a z^{-1} block introduces a unit time _____.

10. If one were to pass a signal through a z^{-m} block, the input signal will be time shifted by _____ delays.

11. In a physical sense, z^{-1} is a _____ element in the digital processor and z^{-m} means that there are _____ elements in the digital processor.

12. Digital filters are broadly classified as _____ filter and _____ filters.

13. FIR filter has only _____-forward gains.

14. All the poles of the FIR filter are located at the _____ of the z-plane.

15. FIR filter is an _____-loop filter.

16. In a FIR filter of order m, for a unit impulse input, the output response will exist upto _____ the sample time.

17. In an IIR filter of order m, for a unit impulse input, the output response will exist upto _____.

18. IIR filter is a _____-loop filter. It contains both the _____ gains and _____ gains.

19. The origin of the s-plane maps to the point _____ in the z-plane.

20. If the poles lie within the _____ circle in the z-plane then the system is stable.

21. The region outside the unit circle in the z-plane is the _____ region.

22. A continuous time signal can be converted into a discrete time signal by _____.

23. The sampling frequency must be greater than _____ the highest frequency component of the analog signal.
24. In an _____ system, a step input will disturb the output response which will grow in amplitude.
25. In a _____ system, a step input will disturb the output response which will just oscillate.
26. In a _____ system, a step input will disturb the output response which will exponentially decay to zero.

DESCRIPTIVE QUESTIONS

1. Why is there a shift from the analog domain control to digital domain control?
2. How are computation, repeatability and re-usability aspects better in the digital domain vis-à-vis the analog domain?
3. For controller design, it is essential that all the system blocks are equivalently either in the analog domain or the digital domain. Based on this what are the two possible approaches and which is the preferred method from the performance point of view?
4. What are the common domains in which a system is usually represented?
5. What is z-transform of a discrete signal?
6. "z^{-1} *block introduces a unit time delay to the input stream*". Discuss.
7. Discuss the FIR and IIR filters using the unit delay blocks.
8. Discuss the canonical or standard form of representation of the IIR filter.
9. Discuss the mapping of the various critical points of the s-plane onto the z-plane.
10. Every strip in the left half s-plane of height $2\pi/T$ beyond $+\pi/T$ and $-\pi/T$ maps onto the same unit circle in the z-plane. Why? Explain.
11. What is the effect of sampling on the signal viewed in the time domain, the frequency domain and the pole-zero domain?
12. What is aliasing? Discuss.
13. How is anti-aliasing usually achieved?
14. What are the different methods of converting the continuous domain models to the discrete domain models?
15. Write short notes on the following: (a) pole-zero map, (b) numerical integration methods, (c) hold equivalence and (d) discretizing state equation.
16. Why is hold equivalence generally the most preferred method of converting the continuous domain model to the discrete equivalent?
17. What are the three performance measures for a system to be controlled?
18. Write short notes on the following: (a) stability, (b) steady-state accuracy and (c) speed of response.
19. What are positional error constant, velocity error constant and acceleration error constant?
20. How is stability measured and specified?
21. How is steady-state accuracy measured and specified?
22. How is speed of response measured and specified?
23. What is bandwidth of a system?
24. What is rise time and settling time?

25. Write short notes on the following: (a) peak overshoot, (b) delay time and (c) error band.

26. "*For an ideal noiseless system with no limit on the system's control input value, the controller gain that gives zero error during transients and steady state is infinite*". Explain.

27. What is the effect on the controller gain due to presence of noise in the system?

28. What is control band?

29. What is limit cycle oscillation? Why does it occur?

30. What are the features of the state space method of controller design?

PROBLEMS

1. Obtain the discrete transfer function of the following digital filters and check their stability:
- **a.** $y(k) = 0.7y(k-1) - 0.4y(k-2) + 0.4x(k)$
- **b.** $y(k) = 1.8y(k-1) - y(k-2) + x(k)$
- **c.** $y(k) = y(k-1) + 0.3y(k-2) + 0.1x(k)$

2. Derive the z-transform for the following signals: (a) unit step, (b) unit ramp, (c) exponential decay with time constant τ, (d) sinusoidal function of frequency ω, (e) co-sinusoidal function of frequency ω.

3. Determine the z-transforms for the following continuous domain transfer functions:

- **a.** $G(s) = \dfrac{10}{s(s+1)^2}$
- **b.** $G(s) = \dfrac{5 \cdot s \cdot e^{st}}{s(s+1)(s+2)}$
- **c.** $G(s) = \dfrac{100}{s^2 + 10s + 25}$
- **d.** $G(s) = \dfrac{10(s-1)}{s(s+2)^3}$
- **e.** $G(s) = \dfrac{10}{s^2 - 5s}$

4. Show that the z-plane variable "z", the s-plane variable "s" and the sampling time T are related by $z = e^{sT}$.

5. The mapping between the s-plane and the z-plane variables is given as $z = e^{sT}$. However, the mappings that result from the numerical integration methods are different from the actual mapping relation given above. How does the approximate mapping relation affect the mapping of the critical points of the s-plane onto the z-plane for the following integration methods: (a) Euler's forward rule, (b) Euler's backward rule and (c) Tustin's trapezoidal rule.

6. Map the pole and zero locations of the following discrete domain transfer functions to the s-plane using (a) exact mapping, (b) Euler's forward rule, (c) Euler's backward rule and (d) Tustin's trapezoidal rule:

- **a.** $H(z) = \dfrac{z(z+1)(z+2)}{(z-0.3)(z^2 - 0.5z + 0.5)}$
- **b.** $H(z) = \dfrac{(z-2)^2}{(z+0.3)(z+0.5)(z-0.7)}$
- **c.** $H(z) = \dfrac{z(z-1)}{z^2 - 3z + 1}$

7. A continuous domain system is given as

$$H(s) = \frac{1}{s^2(s+4)}$$

This continuous system is sampled by means of a sampler followed by a zero-order hold device. This discrete system is in a closed-loop unity

feedback configuration. The controller gain is K. Then

a. What is the discrete transfer function of the open-loop system?

b. What is the discrete transfer function of the closed-loop system?

c. Is the system stable at $K = 1$?

8. A type 1 system has a transfer function given as

$$H(s) = \frac{5}{s(s+4)}$$

This system is connected in a closed-loop unity feedback configuration. Find the positional, velocity and acceleration error constants for the closed-loop transfer function.

9. For the system of Problem 8, determine the steady-state error resulting from the reference input of a ramp waveform which is represented as $r(t) = 5t$.

10. For the system of Problem 8, determine the steady-state error resulting from the input that is given as $r(t) = 5 + 6t + 2t^2$.

ANSWERS

Fill in the Blanks

1. digital; continuous; analog
2. time
3. time
4. frequency
5. pole-zero
6. unit impulses
7. n-domain; z-domain
8. identical
9. delay
10. m time
11. memory; m memory
12. finite impulse response; infinite impulse response
13. feed
14. origin
15. open
16. m times
17. infinity
18. close; feed-forward; feedback
19. $z = 1$
20. unit
21. unstable
22. sampling
23. twice
24. unstable
25. marginally stable
26. stable

CHAPTER 10

Digital Controller Design

Learning Objectives

After reading this chapter, you will be able to:

- design digital controllers using transfer function methods like Bode diagram and root locus plots.
- design digital controllers using state space methods.
- design estimators using state space methods.

This chapter uses the concepts discussed in the previous two chapters to design controllers for systems and more specifically power electronic systems. As majority of the controllers today are in the discrete domain, this chapter primarily addresses the design of digital controllers. Insights into the digital domain design concepts are explained with the help of examples of power electronic systems.

Most literature on control systems devotes considerable space to the stability issues and the methods for finding out the stability and relative stability of the systems. Methods such as Routh–Hurwitz criterion, Jury test, etc. are discussed at length in literature. However, with a mathematical model of the system available, there are a host of programing environments such as SciLAB and MATLAB that will compute the exact location of the poles and zeros of a system of any given order. Therefore, this text does not focus much on the classical methods of stability determination. Instead more space is devoted towards design and examples that will enhance the insights into the system.

The controller design addresses the three performance criteria: (a) stability, (b) steady-state accuracy, (c) speed of response, as discussed in the previous chapter. These three criteria form the specifications for which the controller should be designed. In Section 9.8 of Chapter 9, the general structure of a closed-loop system is discussed. The main function of the controller is to make its input zero by appropriately adjusting its output. This implies a controller with infinite gain. However, as discussed in the above-mentioned section, there are two major issues with infinite gain as controller: (a) The noise problem – every component generates noise and this will also get amplified by the high-gain controller making the output noisy and (b) finite signal limits – the output of the controller and the system control input have finite upper and lower limits. This will cause the infinite gain controller output to clamp to the permissible extreme limits leading to limit cycle saturation operation, which is also called limit cycle oscillation. Both these problems will cause the system to malfunction and the three performance criteria mentioned above may not be met. This chapter will discuss the ways and means of meeting the above-mentioned performance criteria in the presence of noise or external disturbance and with finite control input values.

10.1 Controller Design Techniques

The control of power electronic systems can be visualized by the block schematic shown in Figure 10.1. The controlled plant is a power electronic system wherein either the output voltage or the input current or some specific plant parameter needs to be controlled. The parameter to be controlled is sensed and fed back to be compared with the desired reference after being appropriately processed by a signal processing circuit. The error between the reference and the fed back signal is given to the controller which will appropriately generate a control voltage. For most switched power systems the plant is controlled by duty cycle variation. Therefore a voltage-to-time converter is used to map the controller output to the controller input, which is invariably the duty cycle in switched power systems.

For the system to meet the performance requirements of stability, steady-state accuracy and speed of response, numerous design methods are available for the controller design. The design methods can be broadly classified into the following:

1. **Transfer Function Based Controller Design Methods:** Here the system is modeled primarily as a linear time-invariant (LTI) system with single-input and single–output (SISO). The transfer function along with block diagram algebra is eminently applicable in these methods.
2. **State Space Controller Design Methods:** Here the system is modeled as a general multiple-input and multiple-output (MIMO) plant. The state equation representation of the system is used in these methods.

The traditional and classical techniques are based on the transfer function models of the system. These methods are:

1. Nyquist diagram method.
2. Inverse polar plot method.
3. Nichols chart method.
4. Bode diagram method.
5. Root locus method.

The modern techniques are based on the state equation models of the system. These methods are:

1. full-state feedback;
2. full-state feedback with estimator;
3. output feedback;
4. optimal controller – linear quadratic;

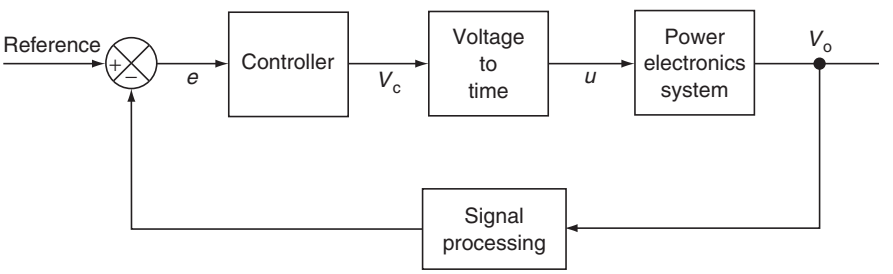


Figure 10.1 | Block diagram of the control scheme of a typical power electronic system.

5. optimal controller with optimal estimator;
6. robust controller;
7. self-tuning and adaptive controllers.

Among the classical methods, though any of the above methods can be used for low-order systems, for high-order systems, the root locus method would in general be a more appropriate approach. The Bode diagram approach can also be used; however, that approach cannot be used for non-minimum phase systems as will be explained later. Calculation of the root loci and the roots of the characteristic equations was a tedious and cumbersome manual process before the advent of the computational tools. Now with software such as SciLAB, MATLAB and other similar computational tools available, plotting the root loci is no longer an issue for complex systems. Therefore, the root locus method itself would be sufficient to design any LTI SISO system of any order by the classical method. However, as Bode diagram is also a popular tool in use by the designers; both these methods will be discussed in the following sub-sections.

The state space methods will also be discussed in this section. The full-state and output feedback methods are discussed in this chapter; however, the optimal and robust controller design methods will be discussed in the next chapter. The self-tuning and adaptive controllers are beyond the scope of this book and are therefore not discussed.

10.2 Bode Diagram Method

This method is a popular method that is used by the practicing engineers as it is relatively easy to check the anticipated results by just obtaining the open-loop frequency response of the system. Bode's contribution to the field of control is in the form of two important statements called Bode's theorems. These two theorems form the basis for the Bode diagram method.

Bode's Theorem I

Statement: The slopes of the asymptotic amplitude–log frequency curve imply a certain corresponding phase shift, and the slope at crossover, where the amplitude–log frequency curve crosses the 0 dB line, is weighted more towards determining system stability than a slope further removed from this frequency.

A slope of 0 dB/decade results in a phase shift of 0, a slope of +20 dB/decade results in a phase shift of +90°, a slope of +40 dB/decade results in a phase shift of +180°, a slope of –20 dB/decade results in a phase shift of –90°, a slope of –40 dB/decade results in a phase shift of –180° and so on. Every slope of +20 dB/decade adds a phase shift of 90° and every slope of −20 dB/decade adds a phase shift of −90°. Further, the slope at the gain crossover frequency is a measure of the relative importance towards determining the system stability. Specifically, at the gain crossover frequency, the phase shift at this frequency is measured to determine the phase margin with respect to 180°.

If a system has a slope of –20 dB/decade at the gain crossover and the other slope sections are far away from the gain crossover frequency, then the phase shift at the gain crossover frequency is approximately 90° and the corresponding phase margin is 90°. Such a value of phase margin implies a stable system. However, if the slope at the gain crossover frequency is –40 dB/decade and the other slope sections are far away from the gain crossover frequency, then the phase shift is approximately 180° and the corresponding phase margin is 0°. This implies that such a system will be on the verge of instability. If the slopes are steeper at the gain crossover frequency, the phase margins would be negative and indicate unstable systems. *Therefore, by Bode diagram approach, it is important to maintain the slope of the amplitude–log frequency curve in the vicinity of the gain crossover frequency at a slope of –20 dB/decade*.

Bode's Theorem II

Statement: *The amplitude and phase characteristics of linear, minimum phase shift systems are uniquely related.*

This means that when the slope of the amplitude–log frequency curve over a particular frequency interval is specified, the corresponding phase shift characteristics over that frequency interval are automatically defined. Conversely, if the phase shift over a particular frequency interval is specified, the corresponding amplitude–log frequency characteristics over that frequency interval are automatically defined.

Bode's second theorem may appear trivial at the outset. However, its implications are quite significant. It should, however, be noted that the second theorem implies that Bode's method should be used only for minimum phase systems because they have minimum phase shift possible for the number of energy-storage elements in the system. This restricts the poles and zeros of the systems to lie on the left half of the *s*-plane or within the unit circle in the *z*-plane. For example, a pole on the left half of the *s*-plane contributes a slope of –20 dB/decade and a phase shift of –90°. A zero on the left half of the *s*-plane contributes a slope of +20 dB/decade and a phase shift of +90°. However, a zero on the right half of the *s*-plane contributes a slope of +20 dB/decade and a phase shift of –90°. Therefore, in such a case even though the system is stable, specifying the phase shift will not define the amplitude–log frequency characteristics and vice-versa. *Therefore, Bode diagram method should not be applied for non-minimum phase systems.*

Bode Diagrams

For a system that has the transfer function model given as $G(s)$, as the *s*-plane variable "*s*" traverses the imaginary axis, it takes the value of the real frequency ω and the plots corresponding to the amplitude and phase angle of $G(j\omega)$ are obtained. The amplitude and the phase angle that are plotted with ω as the common axis are called the Bode diagrams.

The amplitude or magnitude of the transfer function is expressed in decibels (i.e., dB) and the phase angle in degrees. The magnitude M is given as

$$M = 20 \cdot \log_{10} G(j\omega) \tag{10.1}$$

and the phase angle is

$$\phi(j\omega) = \angle G(j\omega) = \tan^{-1}\left(\frac{\text{Imag part } G(j\omega)}{\text{Real part } G(j\omega)}\right) \tag{10.2}$$

Constant

For a transfer function $G(s) = K$, a constant

$$M = 20 \cdot \log_{10} K \tag{10.3}$$

$$\phi(j\omega) = \begin{cases} 0^\circ & K \geq 1 \\ 180^\circ & K < 1 \end{cases} \tag{10.4}$$

Figure 10.2(a) shows the Bode diagrams for the constant transfer function.

Integrator

The transfer function of an integrator is given as $G(s) = 1/s$. Therefore

$$G(j\omega) = \frac{1}{j\omega}$$

The magnitude is given as

$$M = 20 \cdot \log_{10} \frac{1}{\omega} = -20 \cdot \log_{10} \omega \tag{10.5}$$

The phase angle is given

$$\phi(j\omega) = \tan^{-1}\left(-\frac{1/\omega}{0}\right) = -90° \tag{10.6}$$

Figure 10.2(b) shows the Bode diagrams for the integrator. It can be seen from Eq. (10.5) and Figure 10.2(b) that the magnitude plot has a slope of –20 dB/decade and the corresponding phase shift angle is –90°.

Pure Differentiator

Pure differentiator does not exist in nature. However, the transfer function is given as $G(s) = s$ and $G(j\omega) = j\omega$. The magnitude is given as

$$M = 20 \cdot \log_{10} \omega \tag{10.7}$$

and the phase angle is

$$\phi(j\omega) = \tan^{-1}\left(\frac{\omega}{0}\right) = 90^{o} \tag{10.8}$$

Figure 10.2(c) illustrates the Bode diagrams for the differentiator.

First-Order Low-Pass Filter

The transfer function is given as $G(s) = a/(s+a)$ and $G(j\omega) = a/(j\omega + a)$. The magnitude is

$$M = 20 \cdot \log_{10}\left(\frac{a}{j\omega + a}\right) \tag{10.9}$$

and the phase is

$$\phi(j\omega) = \angle G(j\omega) = \angle\left(\frac{a}{a+j\omega} \cdot \frac{a-j\omega}{a-j\omega}\right) = \angle\left(\frac{a^2}{a^2+\omega^2} - j\frac{a\omega}{a^2+\omega^2}\right) = -\tan^{-1}\left(\frac{\omega}{a}\right) \tag{10.10}$$

Figure 10.2(d) illustrates the Bode diagrams for the first-order low-pass filter. The dotted line indicates the exact magnitude characteristics and the dark line indicates the asymptotic approximation. They would differ at most by 3 dB at $\omega = a$. The asymptotic approximations are generally preferred by the designers to quickly obtain the frequency character of the system. For example, in the case of the low-pass filter discussed above, for frequencies much less than "a", the magnitude is 0 dB and for frequencies much greater than "a", the magnitude is a function of ω. Thus,

$$M = \begin{cases} 20 \cdot \log_{10}(1) = 0 & \omega \ll a \\ -20 \cdot \log_{10}(\omega / a) & \omega \gg a \end{cases}$$

The phase shift is as given by Eq. (10.10). At the break frequency where the slope changes, the phase shift is –45°. The first-order low-pass filter can be realized with a resistor and a capacitor. There is only one energy-storing element in the network and it can be seen that the phase shift is the minimum swing between 0° and –90°. This is, therefore, a minimum phase network.

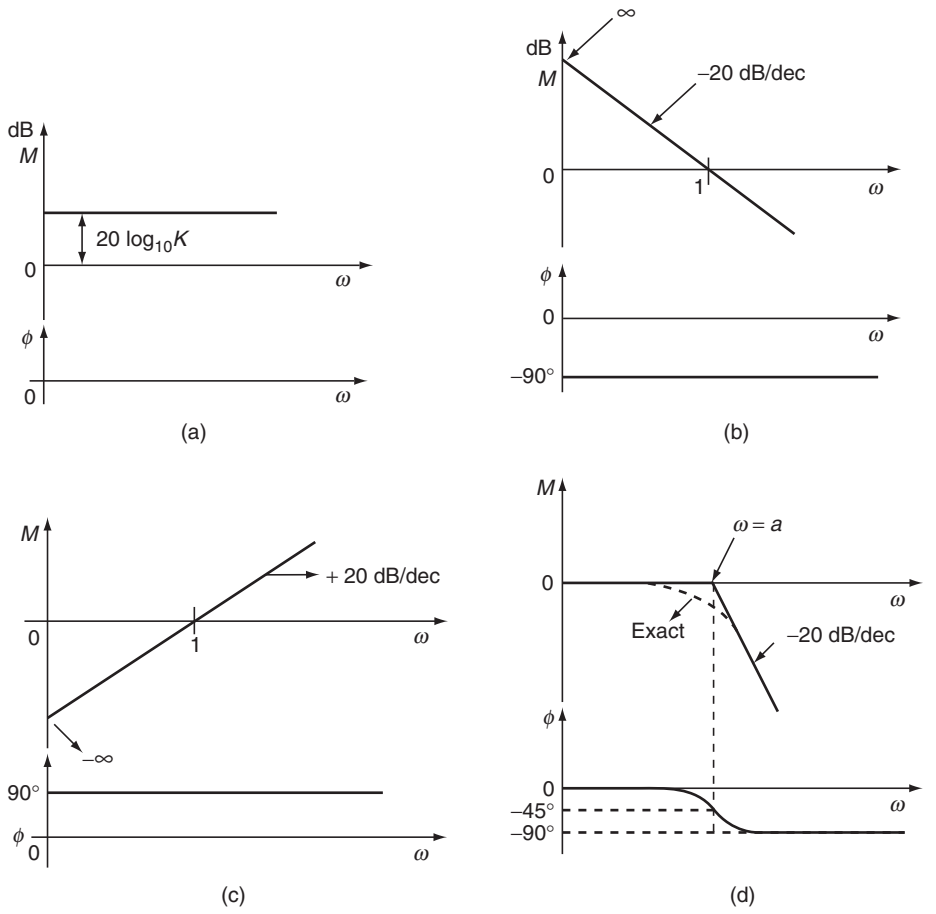


Figure 10.2 Bode diagrams for (a) constant; (b) pure integrator; (c) pure differentiator; (d) first-order low-pass filter.

Lead–Lag Network

The transfer function is given as $G(s) = (s + a)/(s + b)$ and $G(j\omega) = (j\omega + a)/(j\omega + b)$. The magnitude and phase angle are given as

$$M = 20 \cdot \log_{10} \frac{j\omega + a}{j\omega + b} \tag{10.11}$$

$$\begin{aligned} \phi(j\omega) = \angle G(j\omega) &= \angle \left(\frac{a + j\omega}{b + j\omega} \cdot \frac{b - j\omega}{b - j\omega} \right) \\ &= \angle \left(\frac{\omega^2 + ab}{b^2 + \omega^2} - j \frac{\omega(a - b)}{b^2 + \omega^2} \right) = -\tan^{-1} \left(\frac{\omega(a - b)}{\omega^2 + ab} \right) \end{aligned} \tag{10.12}$$

To obtain the asymptotic approximation, consider $a > b$. Then from $\omega = 0$ to $\omega = b$, the magnitude is a constant with amplitude $M = 20 \cdot \log_{10}(a / b)$ with a magnitude slope of 0 dB/decade. From $\omega = b$ to $\omega = a$, the magnitude slope changes by –20 dB/decade with respect to the previous slope due to introduction of the pole or lag effect at $\omega = b$. The phase angle at $\omega = b$ is –45° and drifts towards –90°. From $\omega = a$ onwards, the magnitude slope changes by +20 dB/decade with respect to the previous slope due to the introduction of zero or lead effect at this point. The phase angle is back again at –45° at $\omega = a$ due to positive angle addition to the phase because of the zero. The magnitude slope now becomes 0 dB/decade.

The Bode diagrams for the lead–lag network are illustrated in Figure 10.3. It should be noted that plotting the Bode diagrams is no longer done manually. With the availability of a host of computer programs like SciLAB and MATLAB, the exact bode diagrams are plotted and used for design purposes.

The Bode diagram for any general minimum phase system is shown in Figure 10.4. There are two important parameters that are used in the analysis and synthesis: (a) gain margin and (b) phase margin. In any feedback control system, the feedback is negative. If the controlled system introduces a phase shift of 180°, then along with the negative feedback, the overall phase shift around the loop becomes 360° and therefore positive feedback. If the system gain is greater than 1, the positive feedback will make the output grow till the output saturates or goes into limit cycle oscillation. Therefore, the critical points on the open-loop Bode diagrams are as follows:

1. At the gain crossover frequency, the magnitude $M = 0$ dB or a gain of 1. At this frequency, the phase shift should be within ±180° for the closed-loop system to be stable. The margin that the phase angle has with respect to 180° (or instability) is called the *phase margin*. The phase margin is measured from the ±180° phase angle line to the phase angle curve as shown in Figure 10.4. It is considered positive if the phase margin arrow head is towards the 0° phase line.
2. At the frequency where the phase angle is ±180°, the gain should be less than 1 (i.e., attenuation) or –ve dB such that the system output will not grow. The margin that the gain or magnitude has with respect to 0 dB or unity gain (or instability) is called the *gain margin*. The gain margin is measured from the magnitude curve to the 0 dB line. It is considered positive if the arrow head points upwards.

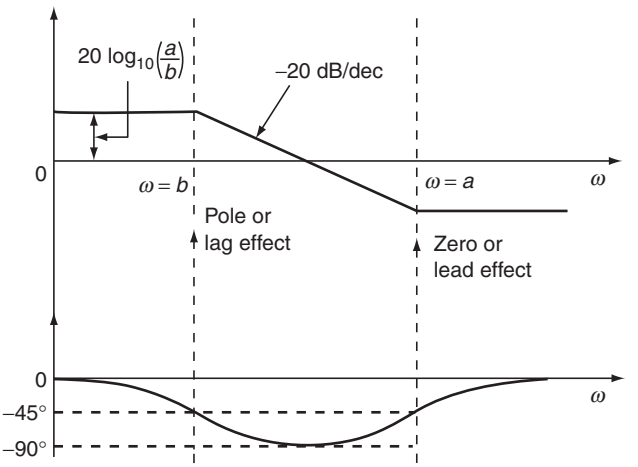


Figure 10.3 | Bode diagram for lead–lag network.

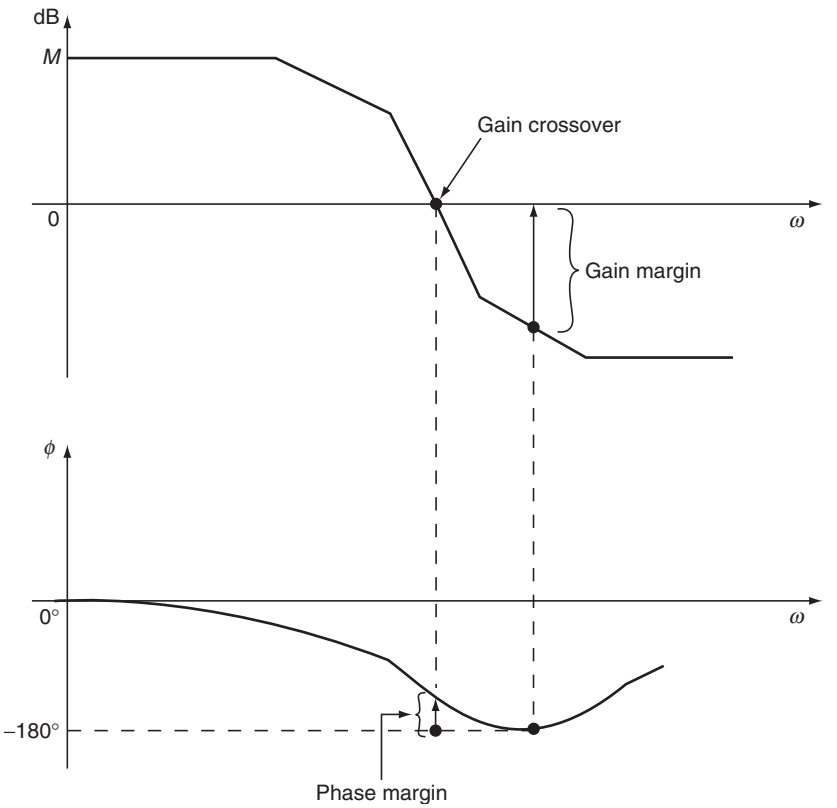


Figure 10.4 | Bode diagram indicating gain margin and phase margin.

Both gain and phase margins are measures of relative stability. The gain towards $\omega = 0$ frequency region gives a measure of the steady-state accuracy. The steady-state portion of the dynamics is near the zero frequency region and the transient portion of the dynamics is the high-frequency region. For the steady-state error to be zero, the gain near the steady-state region should be close to infinity so that the output parameter by the gain will result in almost zero error.

Controller Design with Bode Diagram

The controller design using the Bode diagrams is based on the following three constraints:

1. Gain margin should be positive (e.g., 6 dB gain margin).
2. Phase margin should be positive (e.g., 45° phase margin).
3. The gain or magnitude near the zero-frequency regions should be high to achieve low steady-state error.

The steps for designing the controller are as follows:

Step 1: Plot the amplitude–log frequency and the phase–log frequency Bode diagrams for the system with a specific open-loop transfer function.

Step 2: Find out the gain margin and the phase margin for the open-loop Bode plots.

Step 3: Find out the magnitude near the zero-frequency regions.

Step 4: If the gain margin, phase margin and the low-frequency gains do not satisfy the specifications, then the open-loop Bode plot curves should be shaped accordingly. The shaping is done by including the controller or compensator which will alter the shape of the plant's open-loop Bode diagram. As the controller and the plant are in series, the controller Bode plot will add onto the plant Bode plot to result in the modified Bode plot. This shaping is done based on the following notions:

1. Including a gain in the controller will shift the Bode diagram up or down without affecting the phase angle plot.
2. Including a pole or lag in the controller will add a –20 dB/decade slope to the magnitude curve from the frequency at which the pole or lag is inserted. The phase angle curve from the frequency of pole or lag insertion will shift by at most an additional –90° as given by the lag phase angle.
3. Including a zero or lead in the controller will add a +20 dB/decade slope to the magnitude curve from the frequency at which the zero or lead is inserted. The phase angle curve from the frequency of lead insertion will shift by at most an additional +90° as given by the lead phase angle.
4. The Bode diagram shaping should be done such that at the gain crossover, the slope of the magnitude curve should be –20 dB/decade.

The above steps can be iterated with the help of SciLAB or MATLAB program. Though the Bode diagram has been discussed using the *s*-plane variable, it is applicable without loss of generality to the discrete domain using the *z*-plane variable. In the *s*-plane, the *s*-plane variable traverses the ω-axis to obtain the Bode magnitude and phase diagrams. In the *z*-plane, the ω-axis maps to the circumference of the unit circle. As the *z*-plane variable traverses the circumference of the unit circle, the Bode magnitude and phase diagrams are obtained for the frequency range of $0–\omega_s/2$, where ω_s is the sampling frequency.

EXAMPLE 10.1

Consider the separately excited DC motor that is discussed and modeled in Chapter 8 (Figure 10.5). Let it be required that the speed of the DC motor be controlled.

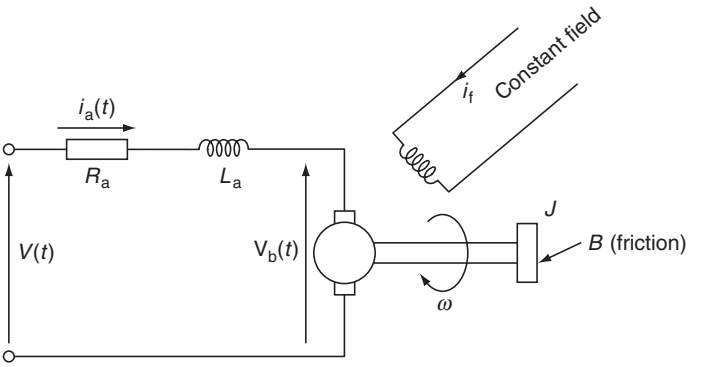


Figure 10.5 Armature-controlled DC servomotor.

Large-Signal Model of Plant

In state space form, the plant model is given by

$$\begin{bmatrix} \mathrm{d}i_a / \mathrm{d}t \\ \mathrm{d}\omega / \mathrm{d}t \end{bmatrix} = \begin{bmatrix} -R / L_a & -K / L_a \\ K / J & -B / J \end{bmatrix} \begin{bmatrix} i_a \\ \omega \end{bmatrix} + \begin{bmatrix} 1 / L_a & 0 \\ 0 & -1 / J \end{bmatrix} \begin{bmatrix} v \\ T_L \end{bmatrix}$$

$$G(s) = \frac{\omega(s)}{V(s)} = \frac{K / L_a J}{s^2 + [(B / J) + (R_a / L_a)]s + [(R_a B / L_a J) + (K^2 / L_a J)]}$$

Plant Parameters

Armature resistance, $R_a = 2\ \Omega$
Armature inductance, $L_a = 0.5$ H
Torque/back emf constant, $K = 0.1$
Friction coefficient, $B = 0.2$ N ms
Load inertia reflected onto the motor shaft, $J = 0.02$ kg m^2/s^2
Sampling time, $T_s = 20$ ms

Transfer Function in the *s*-Domain

The transfer function of the angular frequency with respect to the armature voltage is given as

$$\omega(s) = 10$$

$$\text{----- = ----}$$

$$V(s) = (s + 9.828)\ (s + 4.172)$$

Transfer Function in the *z*-Domain

Conversion from continuous to discrete domain is performed by zero-order hold (ZOH) equivalence.

$$\omega(z) = 0.0018233\ (z + 0.9109)$$

$$\text{----- = ----------}$$

$$V(z) = (z - 0.92)\ (z - 0.8215)$$

State Space Model in the *z*-Domain

Conversion from continuous to discrete domain is performed by ZOH equivalence.

$$A = \begin{bmatrix} 0.92294 & -0.00348 \\ 0.08698 & 0.81856 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.03844 & 0.00182 \\ 0.00182 & -0.90629 \end{bmatrix}$$

$$C = [0\ 1]$$

$$D = [0\ 0]$$
$$x = [ia;\ \omega] \qquad u = [v;\ TL]$$

Bode Plot of Open-Loop Plant

The open-loop Bode plot of the plant is shown in Figure 10.6. It can be observed that

1. The magnitude plot indicates that the gain is never above unity. This implies that the low-frequency or DC gain is very low and therefore the steady-state error is large.
2. Gain margin = 43.3497 dB at –180° phase crossover frequency of 37.131 rad/s.
3. As there is no gain crossover frequency, there is no concept of gain margin here.

To compensate the system, the open-loop Bode plot can be modified by including a compensator or controller such that

1. The DC gain is infinite. This implies that an integrator be used. The integrator scaling factor can be adjusted such that the magnitude curve cuts the 0 dB line so that the phase margin is at least 45° and the gain margin is at least 10 dB.
2. Include a zero in the controller such that the magnitude plot flattens out beyond the zero and the phase shift remains more or less unchanged beyond the zero frequency.

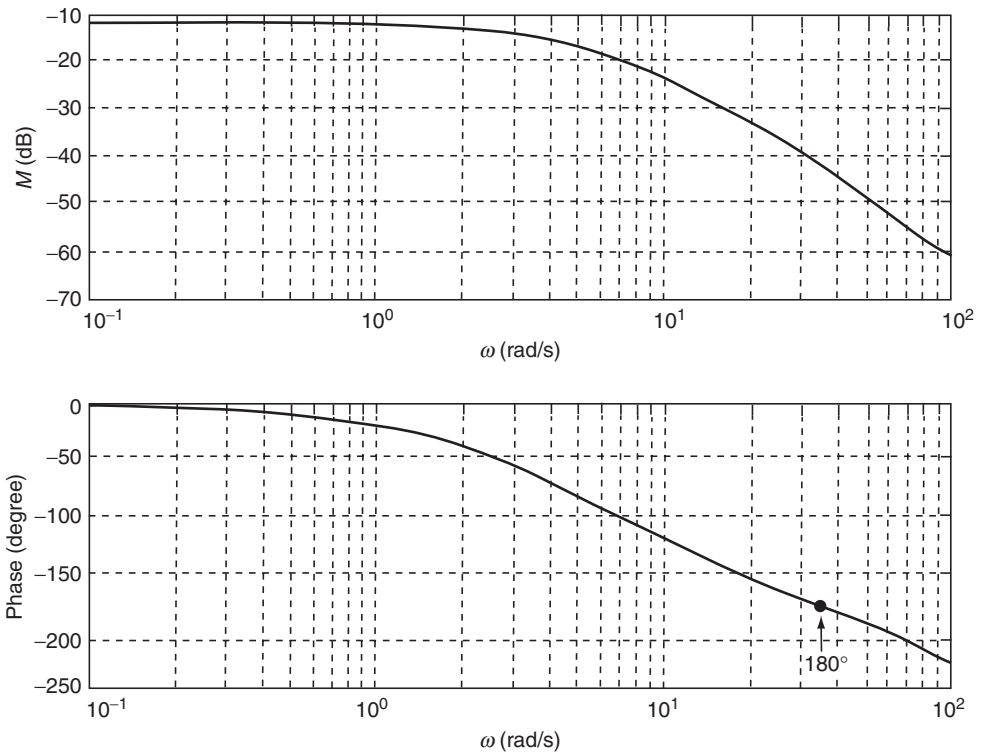


Figure 10.6 Open-loop Bode plot of the plant.

The Bode plot shown in Figure 10.7 is that of the controller or compensator. The magnitude plot shows the presence of an integrator. The inclusion of the zero in the compensator leads to the controller magnitude slope becoming 0 dB/decade. This controller is called a PI controller that is discussed in the next section.
The controller corresponding to the Bode plot of Figure 10.7 is

$$G_c(z) = \frac{6(z-0.9)}{z-1}$$

The Bode plot of the compensated system is shown in Figure 10.8 wherein it is compared with the Bode plot of the uncompensated plant. Observe that the gain margin is about 24 dB and the phase margin is 50.8°. The step response of the closed-loop system is shown in Figure 10.9. The controller and the plant are connected in negative feedback and the equivalent transfer function from the output to the reference input is considered for plotting the closed-loop step response. The step response plot shows that the steady-state error reaches zero. There is a peak overshoot of about 20%. This can further be improved by tuning the waveshape of the compensated system.

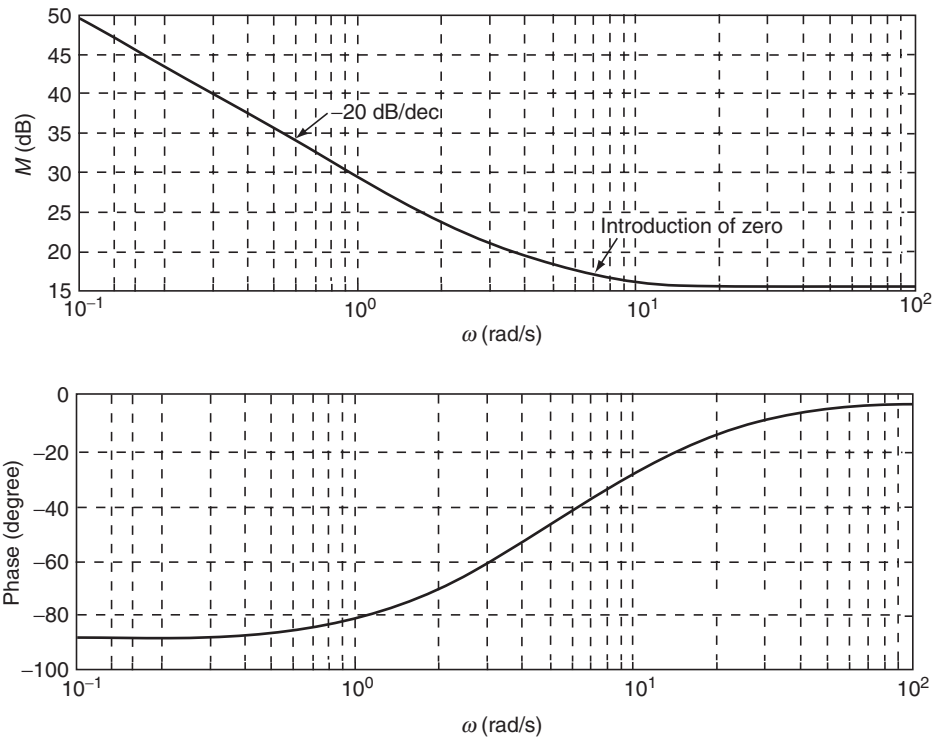


Figure 10.7 | Bode plot of the controller.

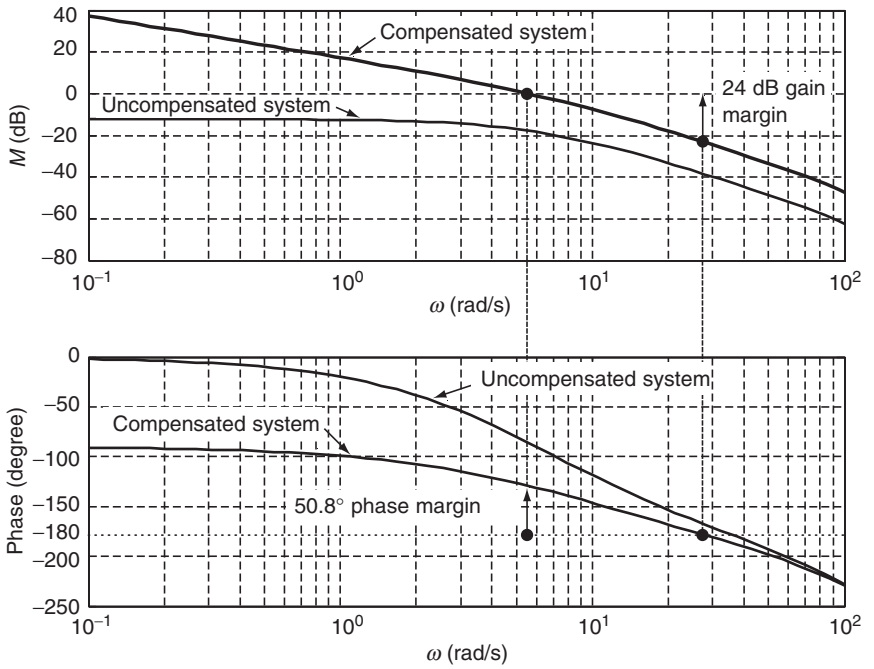


Figure 10.8 | Bode plot of the compensated system.

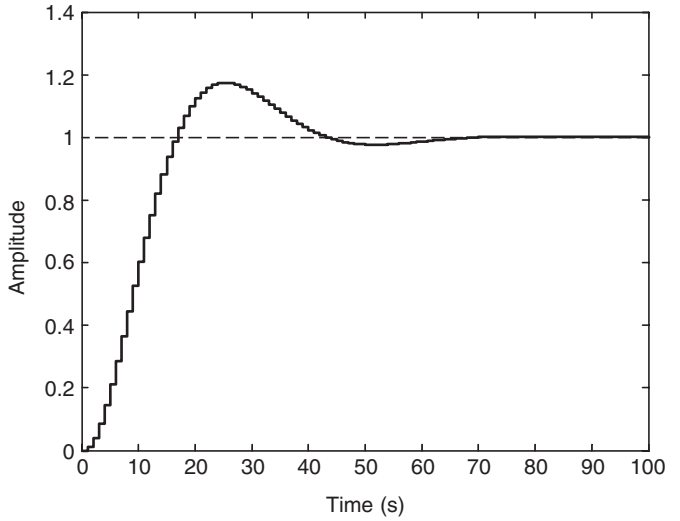


Figure 10.9 | Step response of the closed-loop system.

10.3 PID Controller

PID is an acronym for Proportional (P) – Integral (I) – Derivative (D). The controller with infinite gain would have been the best controller in an ideal sense. If it were not for the noise problem and the finite control input values, the infinite gain controller would have provided zero error under all conditions (either transient or steady state) irrespective of the system that is being controlled. However, in the presence of noise and finite control input limits, the controller cannot be an infinite gain controller as discussed in the previous chapter. Alternately, one can relax the constraints on the controller by stating that instead of providing zero error under all conditions, the controller should at least provide zero error under steady-state conditions. Herein lies the origins of the PID controller. This section will explain the evolution of the PID controller that is probably the most popular controller in the industry today and used in about 85–90% of all control systems.

The noise magnitude is in general a function of the frequency. It is most dominant at higher frequencies. Figure 10.10(a) shows the noise amplitude as a function of the radian frequency. It can be observed that the noise amplitude monotonically increases with frequency. The frequency axis (i.e., the *x*-axis) can be divided into two zones: (a) the low-frequency (LF) zone where the noise is not very significant and (b) the high-frequency (HF) zone where the noise is very dominant. It should, however, be noted that it is difficult to have a very strict line of demarcation between the LF and HF zones. In reality, the transition from the LF to the HF zone and vice-versa is rather hazy. The LF region is associated with the steady-state part of the time domain response and the HF region is associated with the transient part of the time domain response.

Referring to Figure 10.10(a), the controller can have high or infinite gain in the LF zone and high or infinite attenuation in the HF zone. Such a controller assures zero error in the LF region. However, it is difficult or practically impossible to build a controller that swings from $+\infty$ to $-\infty$ as shown (dashed line)

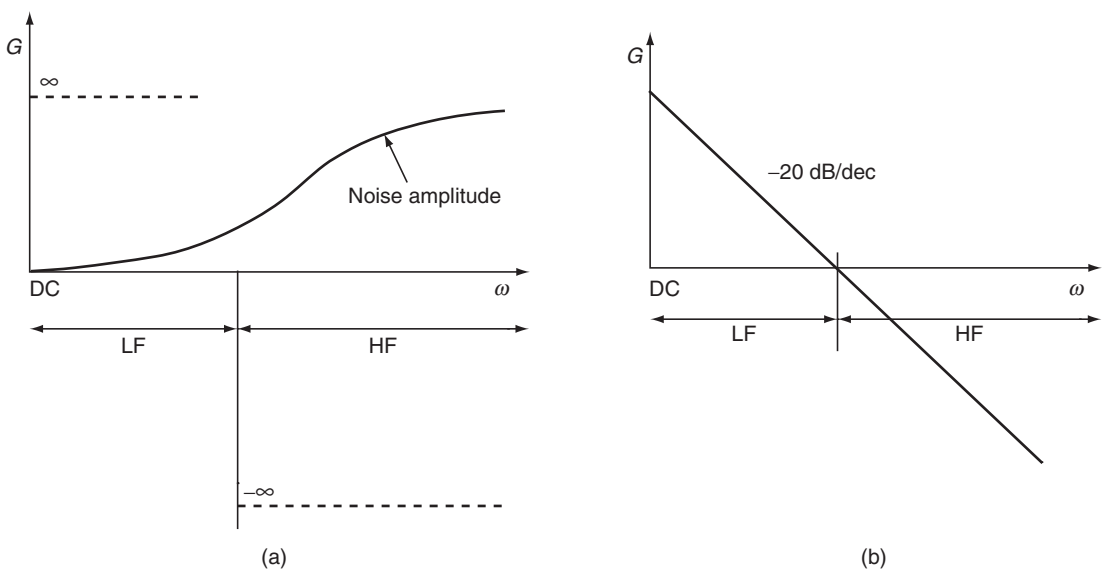


Figure 10.10 (a) Noise amplitude as a function of frequency; (b) integrator gain as a function of frequency.

in Figure 10.10(a). Alternately, one can think of a controller gain that has an infinite gain at DC and the gain gradually decreases with frequency such that in the HF region, the attenuation becomes high as shown in Figure 10.10(b). The simplest block that has a gain profile like that shown in Figure 10.10(b) is that of an integrator. The integrator having a transfer function of $1/s$ has an infinite gain at DC and the gain decreases with frequency at the rate of –20 dB/decade.

More insight into the working of the controller can be gathered by visualizing both the time domain response and the frequency domain gain plot. The LF portion of the frequency domain gain plot is the steady-state portion of the time domain response and the HF portion of the frequency domain gain plot is the transient portion of the time domain response. However, in order to visualize both the plots in a common frame, the time domain response is plotted with the *x*-axis or the time flipped as shown in Figure 10.11(a). This way the LF zone and the HF zones of the frequency domain and time domain response will have reasonable correspondence.

Consider a system that is controlled by a controller having a gain that varies with frequency in a manner as depicted in Figure 10.11(a). The evolution of the output with respect to time is plotted in the time response graph as shown. The time increases from right to left starting from zero. The gain also evolves from right to left on the frequency plot. The "DC" state indicates the steady-state or equilibrium condition of the system.

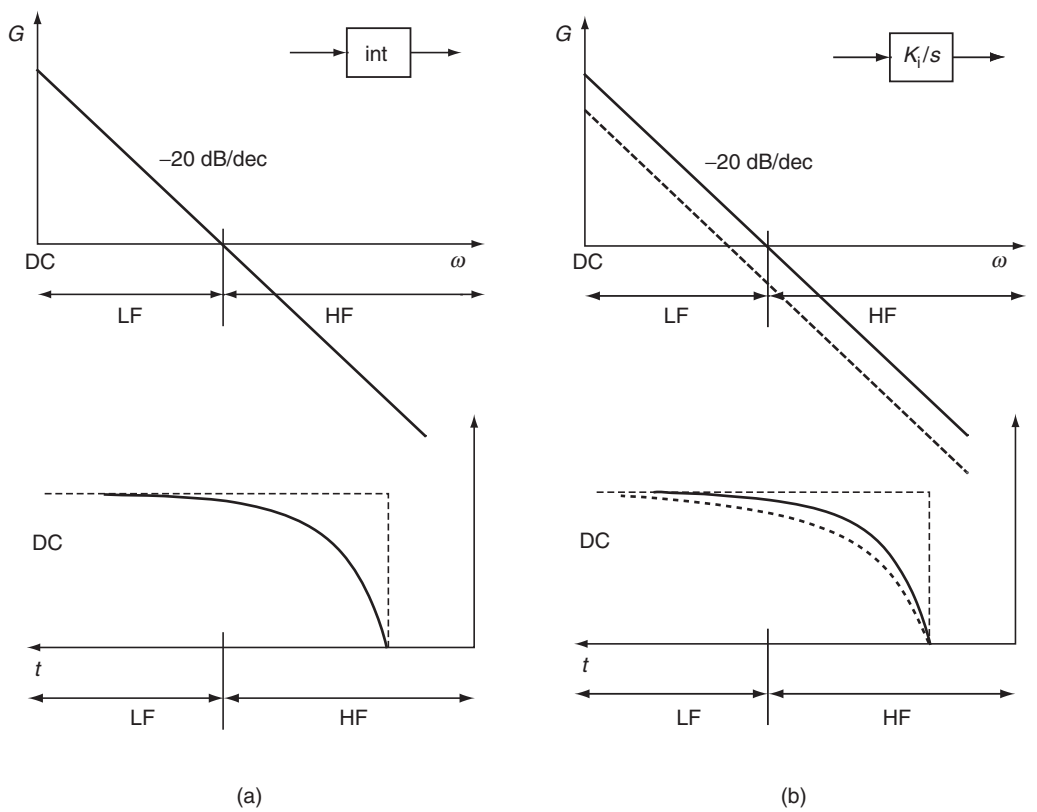


Figure 10.11 (a) Gain versus frequency and time response; (b) scaling the integrator gain.

If a step change is given to the system in terms of either reference change or disturbance, the time frame in the neighborhood of the step change relates to the HF region of the gain plot. During this period when the step change has occurred, the gain is very low and therefore, the error between the output and the reference is very large. As time progresses towards the steady-state or DC conditions, the gain also gradually increases thereby decreasing the error between the output and the reference. At the steady state, the gain is infinite and the error between the output and the reference is zero. This is indicated in Figure 10.11(a). Thus it can be observed that using a simple integrator, one of the most important performance criteria is met, that is, zero steady-state error. Such a controller where only an integrator is used is called *I-control*. Figure 10.11(b) illustrates the effect of the scaling parameter K_i on the integrator gain and the time response. A degree of design freedom can be incorporated into the *I*-controller by introducing a scalar gain scaling factor K_i. The controller transfer function $G_c(s)$ is now given as

$$G_c(s) = \frac{K_i}{s} \tag{10.13}$$

The gain plot is in dB and is given as

$$20 \cdot \log_{10} G_c(\omega) = 20 \cdot \log_{10}(K_i) - 20 \cdot \log_{10}(\omega) \tag{10.14}$$

From Eq. (10.14), it can be observed that at $K_i = 1$, the gain profile is same as that of the pure integrator, that is, $G_c(s) = 1/s$. For values of $K_i > 1$, the gain profile will be higher than the pure integrator gain profile by $20 \cdot \log_{10}(K_i)$ and for values of $K_i < 1$, the gain profile will be less than the pure integrator gain profile by $20 \cdot \log_{10}(K_i)$. This is illustrated in Figure 10.11(b). There are two important points that need to be observed:

1. The value of K_i does not affect the DC gain. The DC gain remains at infinity and hence zero steady-state error is assured for any value of K_i.
2. The value of K_i affects the controller bandwidth. The point of intersection of the gain profile with the 0 dB or the unity gain axis is decided by K_i. This affects the unity gain–bandwidth product of the controller. Therefore, K_i gives a degree of design freedom in the choice of the controller bandwidth or the control speed. If the value of K_i is greater than 1, the speed of the response is faster and if the value of K_i is lesser than 1, the speed of the response is slower as illustrated in Figure 10.11(b).

Figure 10.12(a) shows the next stage in the evolution of the PID controller. The *I*-controller takes care of the steady-state performance. However, the transient performance is poor. To improve the transient performance, the gain during transients, or in other words, the gain in the HF region, should be increased. In the HF region, if a zero is introduced, it will change the slope of the gain profile by adding +20 dB/decade. Thus, the slope becomes zero and the gain profile flattens out as shown. This is achieved by incorporating a scalar gain K_p as shown in the block diagram of Figure 10.12(a). The controller input is passed not only through the I-part, but also through a proportional gain part. The outputs of the two blocks are summed and used as the total controller output. Thus, the overall controller transfer function becomes

$$G_c(s) = \frac{K_i}{s} + K_p = K_p \cdot [\{s + (K_i / K_p)\} / s] \tag{10.15}$$

From Eq. (10.15), it can be seen that a zero is included into the controller by the introduction of the scalar proportional gain K_p in a manner as shown in the block diagram of Figure 10.12(a). This changes the slope of the gain profile from –20 dB/decade to 0 dB/decade from the point on the frequency axis as dictated by the ratio K_i/K_p (i.e., the location of the zero).

It can be observed from Figure 10.12(a) that the gain profile shows an increase in the HF gain with frequency as compared to the pure *I*-controller. This results in a better transient response as indicated in the time response plot.

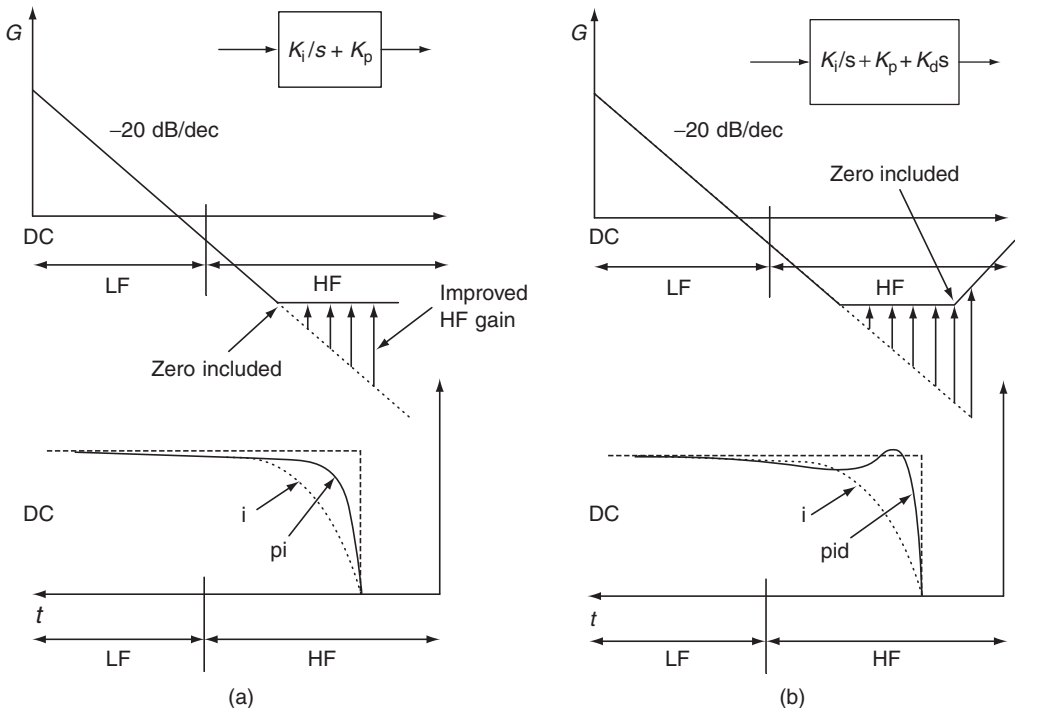


Figure 10.12 | Gain profile and time response for: (a) PI controller; (b) PID controller.

One more degree of freedom can be incorporated by adding a derivative term as shown in the controller block diagram of Figure 10.12(b). Here, the controller transfer function is

$$G_c(s) = (K_i / s) + K_p + K_d \cdot s = K_d \cdot [\{s^2 + (K_p / K_d)s + (K_i / K_d)\} / s] \tag{10.16}$$

From Eq. (10.16), it can be observed that one more zero is added by the derivative term. This zero further adds another +20 dB/decade making the gain profile to increase with a rate of +20 dB/decade as shown in Figure 10.12(b). This further increases the gain improvement with frequency and pulls up the transient response further to make the system fast. However, one must be cautious in introducing the derivative part as this will result in increase in gain in the HF zone which is supposed to have significant noise. Therefore, the scalar K_d is used to scale down the derivative effect such that the system noise does not get amplified. It should also be noted that one cannot in practice build a pure derivative circuit. This will always be associated with an energy-storing element indicating a pole also. Thus, Eq. (10.16) is a first-order transfer function only in theory, but in actual practice it is a second-order transfer function. One order is due to the integrator and the other order is contributed by the non-ideal differentiator circuit or algorithm. However, the pole contributed by the non-ideal differentiator is not a dominant pole and is far away from the origin in the left half of the *s*-plane. Figure 10.13 shows the consolidation of the Bode diagrams of both the plant and the PID controller. Figure 10.13(a) is the magnitude Bode diagram of an arbitrary plant with break frequencies at *a* and *b*. Figure 10.13(b) is the PID controller's magnitude Bode diagram as discussed in this section. Figure 10.13(c) gives the magnitude Bode diagram of the composite system comprising the plant and the

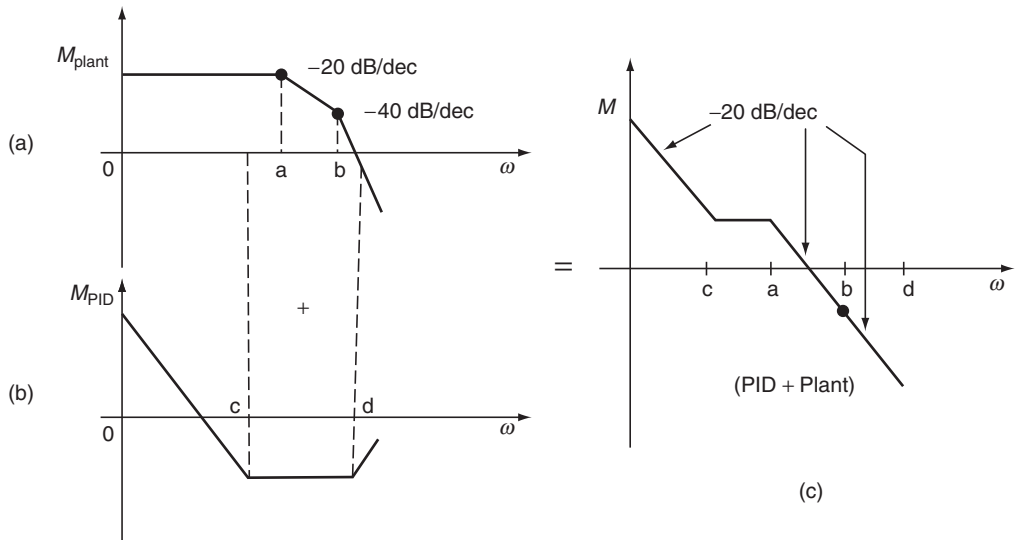


Figure 10.13 | Magnitude plot of: (a) Plant; (b) PID controller; (c) PID controller and plant.

PID controller. The block schematic of the PID controller in a typical closed-loop system is shown in Figure 10.14.

The blocks shown in the PID controller schematic in Figure 10.14 are generic blocks that can be applied for either the continuous systems or the discrete systems. For discrete systems the integrator and differentiator blocks are implemented using the unit delays as explained in Chapter 9.

Figure 10.15 shows the traditional time response of the output of the system to a step change in the reference. Here the time evolves from left to right as is normally the convention. The time response shows the regions where the different parts of the PID controller are active. The I-part or the integral part of the PID controller is active during the steady-state operation wherein the error is controlled to zero.

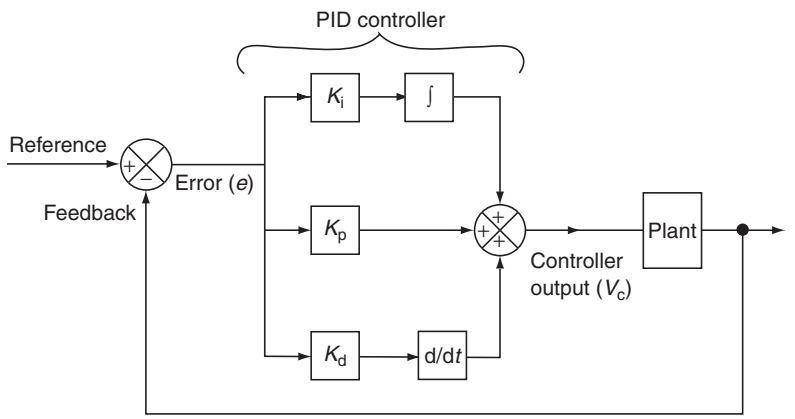


Figure 10.14 | PID controller block schematic.

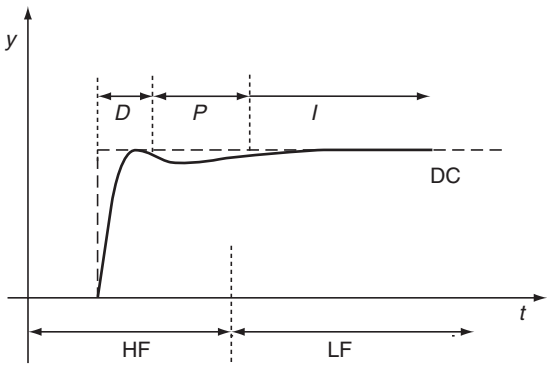


Figure 10.15 | Time response showing the dominant operating regions of the PID controller.

During the transient, the D-part or the derivative part comes into action first as it is more dominant in the HF zone. The D-part will die out quickly and the P-part or the proportional part comes into action to act as a bridge between the D-part and the I-part. It should be noted that if the plant already has an integrator in-built into it or that it has an integrator pole either $s = 0$ or $z = 1$, then K_i should be zero. As the integrator of the plant will provide the necessary infinite gain at steady-state conditions, the external integrator in the controller is not necessary. Therefore, a propotional derivative (PD) controller is sufficient.

PID Controller Design

The PID controller parameters K_i, K_p and K_d can be designed

1. by the Bode diagram controller design method if the transfer function of the plant is known;
2. by experimental method if the model of the plant is not known.

The Bode diagram controller design method has been discussed in the previous sub-section. This sub-section will discuss the experimental method.

The experimental method for determining the parameters of the PID controller is used for plants where the mathematical model of the plant is not known. The experimental setup is as shown in Figure 10.16.

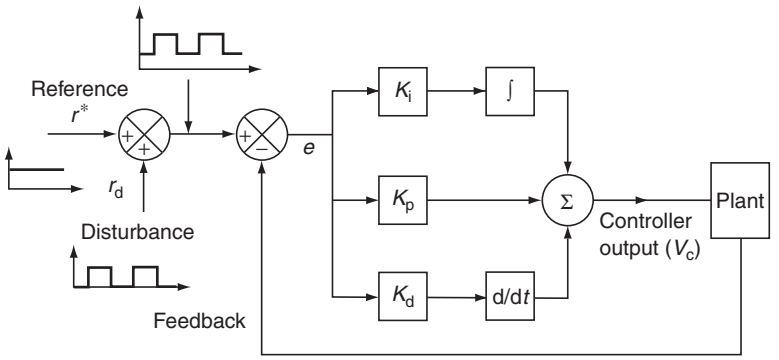


Figure 10.16 | Experimental setup for tuning the parameters of the PID controller.

The reference for the system is now modified to include a disturbance signal r_d. The disturbance signal is a square pulse waveform that has an amplitude that is 10% of the actual reference r^*. Thus the modified reference that is applied to the system is $r^* + r_d$. This modified reference consists of the actual reference r^* superimposed with the square pulse disturbance. The feedback signal along with the modified reference is shown in Figure 10.17.

The steps for experimentally tuning the parameters of the PID controller are as follows:

Step 1: Set $K_p = 0$ and $K_d = 0$.

Step 2: Set the frequency of the superimposed square pulse to 1 s. This should be adjusted such that the period is much greater than the time constant of the system.

Step 3: Observe the feedback signal with respect to the modified reference as shown in Figure 10.17 on an oscilloscope.

Step 4: Start with a very small value of K_i and keep increasing K_i till it overshoots and/or damped oscillations result in the feedback signal. Select the value of K_i to be half the value of K_i that just results in overshoots and/or damped oscillations in the feedback signal.

Step 5: Set K_i to a value as obtained in Step (4). Set $K_d = 0$ and repeat the exercise of Step (4) by gradually increasing K_p from a very small value.

Step 6: Set K_i to a value as obtained in Step (4) and K_p to the value as obtained in Step (5). Gradually increase K_d till satisfactory time response in the transient region is obtained.

Integrator Windup in PID Controller

In some cases, the control input to the plant will have a limiter either built into the plant or included external to it in order to limit the control input signal swing to be within the limits accepted by the interface circuitry. For example, the PID controller may be designed with power supplies of ±15 V and the control input to the plant may be constrained to accept signals only within ±5 V. In such cases, to protect the input circuitry of the plant, the controller output will be passed through a limiter as shown in Figure 10.18. In such situations, due to the memory effect of the integrator in the controller, the integrator winds up and causes a delay in the response of the controller to the changes in the error. *This problem is called integrator windup*.

The integrator windup is illustrated in Figure 10.19. Consider an error signal *e* as shown in Figure 10.19(a). The output of the integrator V_c linearly builds up as shown. The control input *u* follows the controller output till the limiter's upper limit of *u*+ is reached. When V_c reaches *u*+, the control input *u* is clamped to

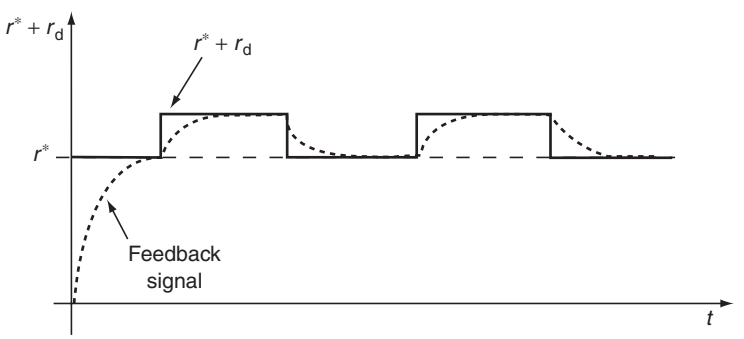


Figure 10.17 | Response of the plant to the square pulse superimposed reference.

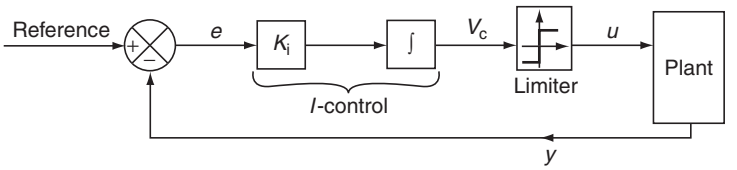


Figure 10.18 Schematic showing the integral portion of the controller (*I*-control) with output limiter.

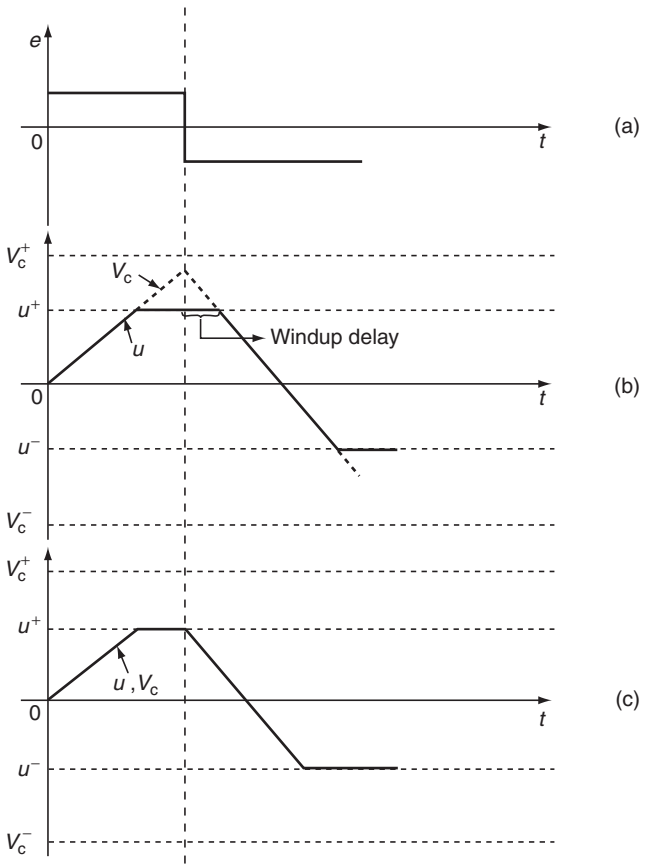


Figure 10.19 (a) Error waveform; (b) integrator output and control input showing windup delay; (c) integrator output and control input with an anti-windup solution.

$u+$ due to limiter action. However, V_c continues to rise as the integrator limit is V_c+. Due to feedback action, if the error e changes as shown in Figure 10.19(a), the integrator de-accumulates and V_c starts decreasing as shown in Figure 10.19(b). However u starts to change and follows the controller output only after V_c drops below $u+$. The time taken from the instant the error signal changes to the instant when the control input u changes to take corrective action, is the delay in the controller response. This is due

to the fact that the integrator output had wound up to a higher value and takes some time to unwind and come within the control band.

This problem is solved by incorporating anti-windup circuitry or anti-windup algorithm in the controller implementation such that the integrator does not integrate beyond the limiter's limit. In other words, the integrator is inhibited from accumulating if the control input has reached its limits for a given error signal polarity. Figure 10.19(c) shows the integrator output and the control input with anti-windup solution incorporated into the integrator. It can be observed that the control input responds immediately to changes in the error signal as the integrator output is prevented from winding up to a higher value than the limits of the limiter.

In the case of PID controllers built with op-amps, back-to-back Zener diodes are connected to clamp the voltage across the capacitor element that acts as the accumulator. This will ensure that the capacitor voltage does not windup. In the case of the discrete implementation, the input to the integrator is controlled through a Boolean logic as follows:

1. For a positive error, if the output of the PID controller is greater than or equal to the control input upper limit, then the integrator input is set to zero.
2. For a negative error, if the output of the PID controller is lesser than or equal to the control input lower limit, then also the integrator input is set to zero.
3. The integrator is allowed to integrate only if the above two conditions are not violated.

Practical PID Controllers

The PID controller transfer function as given by Eq. (10.16) is

$$G_c(s) = (K_i / s) + K_p + K_d \cdot s = K_d[\{s^2 + (K_p / K_d)s + (K_i / K_d)\} / s]$$

However, this is not a practical transfer function as a pure differentiator cannot be implemented. The controller transfer function will also have another pole in addition to the integrator pole to make it a controller that can be implemented. Thus the PID controller transfer function becomes

$$\begin{aligned} G_c(s) &= \frac{K_i}{s} + K_p + K_d \frac{s}{s + s_d} \\ &= \left(\frac{K_p + K_d}{s_d} \right) \left(\frac{s^2 + [(K_i + K_p \cdot s_d) / (K_p + K_d)]s + [(K_i \cdot s_d) / (K_p + K_d)]}{s[(s / s_d) + 1]} \right) \end{aligned} \tag{10.17}$$

where s_d is the differentiator pole frequency. This will change the high-frequency portion of the PID controller Bode diagram as shown in Figure 10.20.

For a discrete implementation, the PID controller transfer function in the z-domain is given as

$$G_c(z) = \frac{K_i}{z-1} + K_p + \frac{K_d}{T} \left(\frac{z+1}{z} \right) \tag{10.18}$$

where the integrator is based on Euler's method and the differentiator is based on the first-order difference equation as discussed in Chapter 9.

The analog implementation of the PID controller circuit using op-amps is shown in Figure 10.21. The op-amp-1 is a difference amplifier that gives the error between the reference and the feedback signal. The error is fed to three sets of op-amps in parallel. Op-amp-2 is the integral controller. Two back-to-back

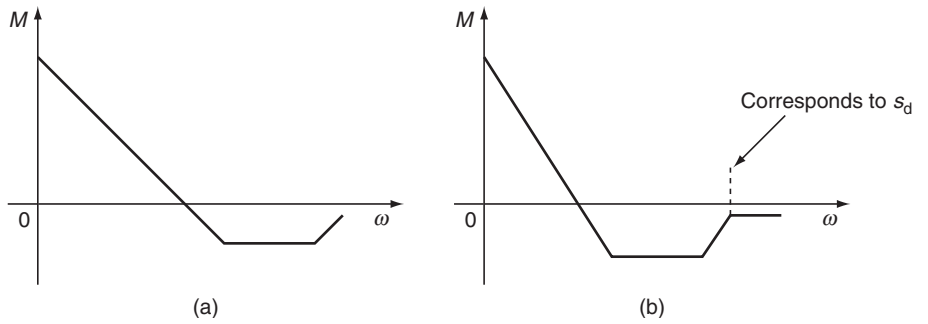


Figure 10.20 | Magnitude plot of (a) an ideal PID; (b) a practical PID with differentiator pole.

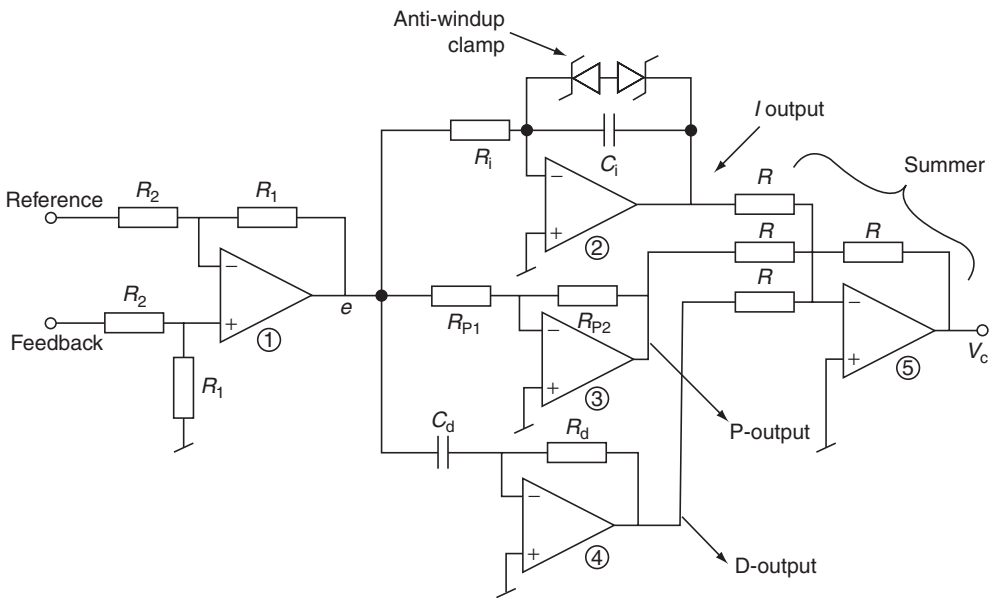


Figure 10.21 | PID circuit schematic.

Zeners are connected across the capacitor C_i to provide a simple clamp to prevent windup of the voltage across C_i. Op-amp-3 is a simple inverting amplifier that provides the proportional gain to the error signal. Op-amp-4 is the differentiator. The outputs of the three stages are summed up in op-amp-5 to provide the overall PID controller output. The transfer function of the circuit is

$$G_c(s) = \frac{V_c(s)}{E(s)} = \frac{1/R_i C_i}{s} + \frac{R_{p2}}{R_{p1}} + R_d C_d s \tag{10.19}$$

One should note that the differentiator is not ideal as indicated in Eq. (10.19). The op-amp acts as a low-pass filter and provides the high-frequency differentiator pole to make the system a second-order system. On comparing Eq. (10.19) with Eqs. (10.16) and (10.17), the relationships between K_i, K_p and K_d

with the circuit parameters can be established. After K_i, K_p and K_d are designed as per the discussion in a previous section, the circuit parameters can be determined by comparing Eqs. (10.16) and (10.19).

Figure 10.22(a) shows another PID circuit that uses only one op-amp. The circuit introduces a 180° phase reversal. Therefore, the error signal is inverted by interchanging the reference signal and the feedback signal in the difference amplifier op-amp circuit. This will give the inverter error signal output that will cancel out the phase reversal introduced by the PID op-amp circuit. The transfer function is given as

$$G_c(s) = \frac{V_c(s)}{E(s)} = \frac{sR_1C_1 + 1}{sC_1R_2} + \left(\frac{sR_1C_1 + 1}{sR_3C_3 + 1}\right)\frac{C_3}{C_1}$$

$$= \left(\frac{R_1R_3C_3 + R_1R_2C_3}{R_2}\right)\frac{s^2 + \left(\frac{R_1C_1 + R_3C_3 + R_2C_3}{R_1C_1R_3C_3 + R_1R_2C_1C_3}\right)s + \left(\frac{1}{R_1C_1R_3C_3 + R_1R_2C_1C_3}\right)}{s(sR_3C_3 + 1)} \quad (10.20)$$

The transfer function given in Eq. (10.20) is of the form given in Eq. (10.17). In many applications, a PI controller alone is sufficient. Figure 10.22(b) gives the circuit schematic of PI controller. The transfer function is given as

$$G_c(s) = \frac{V_c(s)}{E(s)} = \frac{sR_1C_1 + 1}{sC_1R_2} = \frac{1/R_2C_1}{s} + \frac{R_1}{R_2} \quad (10.21)$$

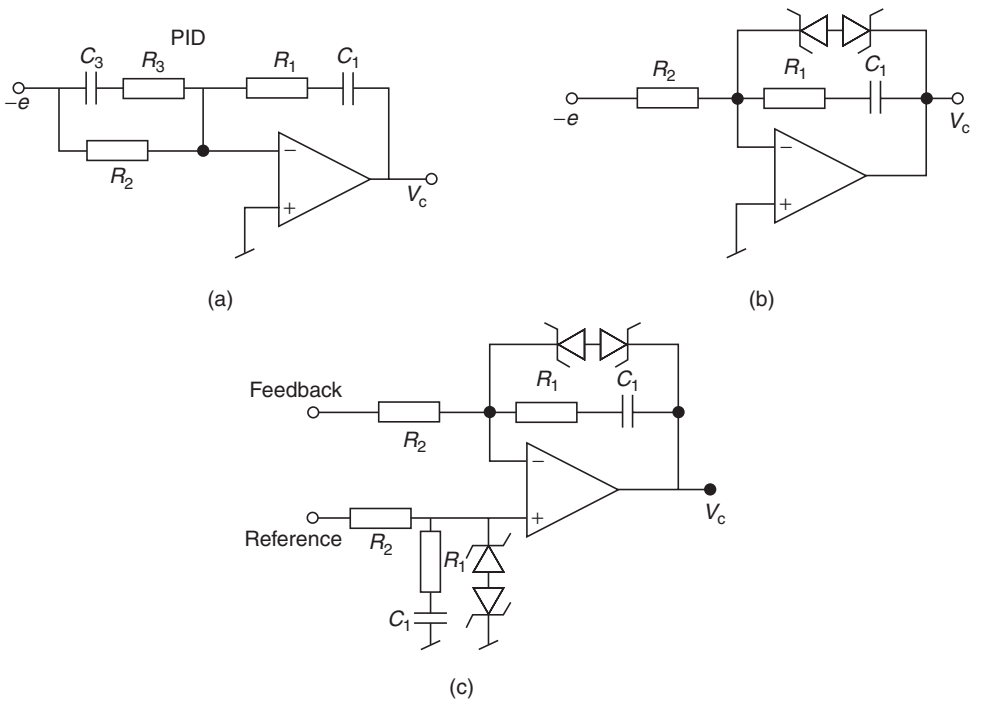


Figure 10.22 (a) Another PID circuit schematic; (b) single op-amp PI circuit schematic; (c) single op-amp comparator and PI circuit schematic.

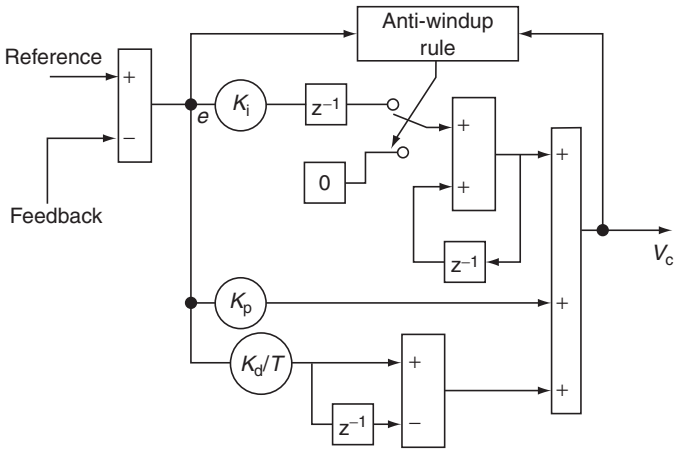


Figure 10.23 | Discrete implementation of PID controller.

This is of the form

$$G_c(s) = \frac{V_c(s)}{E(s)} = \frac{K_i}{s} + K_p$$

where

$$K_i = \frac{1}{R_2 C_1} \text{ and } K_p = \frac{R_1}{R_2}$$

The PI controller of Figure 10.22(c) is interesting in that it uses only one op-amp even eliminating the need for the difference op-amp needed to obtain the error. This acts like a difference amplifier and has the transfer function as given by Eq. (10.21). The disadvantage of this circuit is that two sets of matched R_1, R_2 and C are needed to obtain accurate performance.

The discrete implementation is shown in Figure 10.23. The anti-windup solution for the integrator is implemented by comparing the error signal and the PID output signal. The integrator input is either connected to the error signal or to the zero value in accordance to the logic described in the section "Integrator Windup in PID Controller". Other numerical integration and differentiator algorithms can also be used in place of the simple algorithms used in Figure 10.23 to obtain better controller output accuracy.

10.4 Root Locus Method

The root locus method is a graphical technique for determining the closed-loop poles of a system as a function of the gain. The relationship that exists between the poles of the closed-loop transfer function and the gain, poles and zeros of the open-loop transfer function is the basic principle of this method. The root locus method gives a complete and accurate transient and steady-state picture of the closed-loop system as the closed-loop poles can be directly obtained from the root loci. Plotting the root locus on the *s*-plane or the *z*-plane by manual means is very cumbersome and tedious. Therefore, this method, though complete and detailed, was not very popular with the designers till the advent of computer programs that removed the drudgery and tedium of plotting the root loci. Among the classical techniques for SISO systems, this method by far offers the most complete, detailed and accurate transient and steady-state solution for the system.

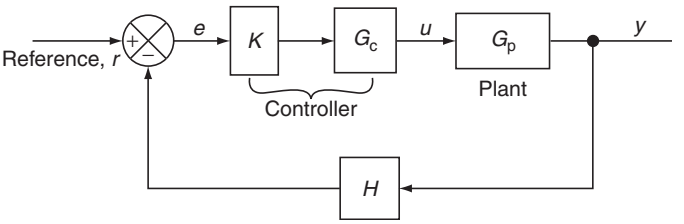


Figure 10.24 | General structure of closed-loop system.

Figure 10.24 shows the general structure of a closed-loop system wherein the plant G_p is an LTI SISO system. The block H is the signal processing circuitry of the feedback signal. The controller is divided into two parts: (a) the gain part K and (b) the part representing the pole–zero structure of the controller, G_c.

Without loss of generality, the closed-loop system can represent either a discrete domain system or a continuous domain system. In a continuous domain system, the plant transfer function is $G_p(s)$, the feedback block transfer function is $H(s)$ and the controller pole–zero structure part is expressed as $G_c(s)$. In the case of the discrete domain system, the continuous domain plant is converted into the discrete equivalent as discussed in Chapter 9 and is expressed as $G_p(z)$, the feedback block is expressed as $H(z)$ and the controller pole–zero structural part is expressed as $G_c(z)$. The discussion to follow is equally applicable to both continuous domain and discrete domain systems. However, as majority of the controllers in the products today are in the discrete domain, majority of the design examples will be in the discrete domain.

Root Locus Plot

Consider the system shown in Figure 10.24. The closed-loop transfer function is given as

$$G_{\text{close}}(z) = \frac{Y(s)}{R(s)} = \frac{K \cdot G_c \cdot G_p}{1 + K \cdot G_c \cdot G_p \cdot H} \tag{10.22}$$

From the closed-loop transfer function, the closed-loop poles are given by the roots of the denominator of Eq. (10.22), that is, the roots of the characteristic equation. Thus the following equation,

$$1 + K \cdot G_c \cdot G_p \cdot H = 0$$

is the characteristic equation. All the components of the characteristic equation are obtainable from the open-loop model of the system. Here G_p is the plant model, H the model of the feedback signal conditioner, G_c the pole–zero structure of the controller that is defined by the designer, K the controller gain. Therefore, to find the poles of the closed-loop transfer function,

$$\begin{aligned} 1 + K \cdot G_c \cdot G_p \cdot H &= 0 \\ K \cdot G_c \cdot G_p \cdot H &= -1 = 1\angle(2n+1)\pi \end{aligned} \tag{10.23}$$

where $n = 0, \pm 1, \pm 2, \ldots$. From Eq. (10.23) the existence of a closed-loop pole at a point on the s-plane or z-plane is obtained if the following two constraints are satisfied:

1. Magnitude constraint, $\left|KG_cG_pH\right| = 1$.
2. Angle constraints, $\angle KG_cG_pH = (2n+1)\pi$.

In the two constraints stated above, at different values of K, different points on the s- or z-plane are obtained that would satisfy the constraints and become eligible to be closed-loop poles. Thus as K is varied from 0 to ∞, all possible closed-loop poles are located on the s- or z-plane for any given K. The locations of the closed-loop poles on the s- or z-plane form lines as K is varied and are called the *loci of the roots of the characteristic equation or the root loci*.

There will be as many root loci as the order of the system. This is due to the fact that there will be as many closed-loop poles as the order of the system and therefore as many root loci. Each segment or branch of the root locus describes the variation of the particular closed-loop pole as the gain is varied.

The open-loop poles define the start of the root locus when $K = 0$ and the open-loop zeros define the termination of the root locus when $K = \infty$. This can be easily concluded from the magnitude condition of Eq. (10.23), that is, $\left|KG_cG_pH\right| = 1$. At the open-loop poles, $\left|G_cG_pH\right|$ is infinite. To satisfy the magnitude constraint, $K = 0$. At the open-loop zeros, $\left|G_cG_pH\right|$ is zero. To satisfy the magnitude constraint, $K = \infty$. When the order of the denominator is greater than the numerator, that is, when there are more open-loop poles than zeros, the number of root loci amounting to the difference between the number of open-loop poles and zeros will terminate at infinity.

Controller Design

With the concepts of the root locus discussed above, a controller may be designed for a given system using the root locus method to meet required performance specifications. In designing a controller for a system based on the root locus method, the following guidelines will help in reducing the design iterations:

1. The root locus method should be performed with the aid of a computer program like SciLAB or MATLAB in order to be effective.
2. The controller structure should be decided first. This is normally done intuitively. Therefore, the controller design for a given system and performance requirement need not be unique.
3. If the controlled plant does not contain an open-loop pole at $s = 0$ or $z = 1$, then the steady-state error will not be zero. An integrator pole should be built into the controller pole–zero structure in such cases. If the controlled plant contains an open-loop pole at $s = 0$ or $z = 1$, then avoid the integrator pole in the controller pole–zero structure. If it is not possible to place an integrator pole in the controller due to unsatisfactory root loci, then choose the closed-loop poles at as large a value of K as possible to reduce the steady-state error.
4. A given system may have unstable poles or right half plane zeros. An example of right half plane zero is the boost converter. An example of a real system having an unstable pole is a magnetic levitation system where the operating point is inherently unstable but has to be maintained by control action. In all such applications the closed-loop poles should be brought inside the unit circle in z-plane or the left of the s-plane by proper selection of the controller pole–zero structure.
5. Poles of the controller cannot be placed outside the unit circle or on the right half of the s-plane.
6. Since the closed performance is determined by the closed-loop poles, to increase the speed of response, the root loci should be shaped to move closer to the origin in the z-plane or away from the negative real axis in the s-plane.
7. As far as possible do not select closed-loop pole locations that are close to the boundary of stability. The gain and other system parameters will vary and may cause the root loci to drift during operation, and as a result the closed-loop pole location at the selected gain K may drift into the unstable region. To design robust controllers, select the closed-loop poles that are far from the boundary of stability (i.e., the unit circle in the case of discrete controller or the imaginary axis in the case of continuous controller).

8. While determining K, it must be remembered that the number of zeros introduced cannot exceed the number of poles introduced. Also, complex poles or zeros will always occur in pairs. This implies that energy is stored in both the potential and the kinetic storage components, resulting in exchange of energy between them. It never reaches equilibrium if there is no dissipation, that is, if the poles are on the imaginary axis. If there is a dissipative element in the system, then the pole is away from the imaginary axis (s-plane) or within the unit circle (z-plane).
9. The pole–zero structure of the controller has to be defined to appropriately shape the root loci. This is the most uncertain part in the whole design process. This will improve with experience. However, the following intuitive insight can be used for selecting the pole–zero structure of the controller to shape the root loci. As a general guideline, zeros attract the root loci and poles repel the loci. Hence to shape a portion of the loci closer to a desired region in the s- or z-plane, zeros may be used. In a similar manner, controller poles should also be placed. It must be remembered that addition of a zero means that you must add a pole also to make it practically implementable. Care should be taken to add just the minimum number of poles and zeros so that the overall order of the system is not unduly increased which will deteriorate the system dynamics.
10. The origin in the z-plane corresponds to negative infinity region in the s-plane. This means that if all closed-loop poles are chosen to be close to these regions, the system response will be faster. This also implies the closed-loop system will be more robust to system parameter variations because the drift in the root loci in the face of variations in the system parameters will not affect the relative stability significantly as the closed-loop poles are far removed from the instability boundary.

Design Steps

Referring to Figure 10.24 and the guidelines discussed above, the controller design procedure using the root locus method is given in the following steps which are algorithmic in nature. A simple program can be written and used to automate the steps. This will help in quick iterations and fine tuning the controller.

Step 1: Model the plant in the continuous time domain. This is based on the discussion in Chapter 8. This will result in the plant transfer function $G_p(s)$.

Step 2: Transform the plant model to the discrete domain using the continuous-to-discrete domain transformation methods discussed in Chapter 9. The ZOH method is the continuous-to-discrete transformation method that fits most applications that use sample-and-hold (S/H) circuits to perform analog-to-digital data conversions. At the end of this step, the discrete domain plant transfer function $G_p(z)$ is obtained. If the controller is to be designed in the continuous domain, this step is skipped. Let the plant transfer function be

$$G_p = \frac{n_p}{d_p} \tag{10.24}$$

where n_p is the numerator polynomial of the plant transfer function that denote the zeros and d_p is the denominator polynomial of the plant transfer function that denote the poles.

Step 3: Obtain the transfer function of the feedback sensing and processing circuitry and transform it from the continuous domain to the discrete domain using the ZOH. Let the feedback block transfer function be

$$H = \frac{n_h}{d_h} \tag{10.25}$$

where n_h is the numerator polynomial denoting the zeros and d_h the denominator polynomial denoting the poles.

Step 4: Define the pole–zero structure for $G_c(z)$ or $G_c(s)$ depending on whether the controller design is in discrete domain or continuous domain, respectively. Let the controller transfer function without the gain be

$$G_c = \frac{n_c}{d_c} \tag{10.26}$$

where n_c is the numerator polynomial of the controller transfer function that denotes the zeros and d_c is the denominator polynomial of the controller transfer function that denotes the poles.

Step 5: Plot the root locus of the loop transfer function G_cG_pH with the controller gain K as the parameter that is varied from 0 to ∞. Root locus plotting functions are available in computer programs like SciLAB and MATLAB. The loop transfer function without K is

$$G_l = \frac{n_c \cdot n_p \cdot n_h}{d_c \cdot d_p \cdot d_h} \tag{10.27}$$

Step 6: Locate the closed-loop poles on the root locus by applying the guidelines discussed above. Note down the value of the controller gain K at the selected closed-loop pole location.

Step 7: With value of gain K, obtain the closed-loop transfer function. The closed-loop transfer function is

$$G_{close} = \frac{K \cdot G_c \cdot G_p}{1 + K \cdot G_c \cdot G_p \cdot H} = \frac{K \cdot n_c \cdot n_p \cdot d_h}{d_c \cdot d_p \cdot d_h + K \cdot n_c \cdot n_p \cdot n_h} \tag{10.28}$$

Step 8: Obtain the step response of the closed-loop transfer function given by Eq. (10.28). The step response plotting functions are available in computer programs like SciLAB and MATLAB.

Step 9: Observe the step response to check if the performance specifications are met.

Step 10: If performance specifications are not met, Steps (4)–(9) should be iterated till the performance is satisfactory. In performing the iteration, first iterate by changing the locations of the poles and zeros of the controller without increasing the order of the system. Only if it is not possible with this pole–zero structure, should one change the order of the controller to obtain satisfactory response.

In addition to the guidelines discussed above, following are some tips and hints for defining the pole–zero structure of the controller for achieving suitable root loci:

1. All open-loop poles and zeros are within unit circle:
   - If there is no pole at $z = 1$, place 1 there for zero steady-state error.
   - Place a zero such that PI controller structure is obtained.
2. All poles are within the unit circle (z-plane) and some zeros are outside it or all poles are in the left half of the s-plane and some zeros are on the right half:
   - As the closed-loop poles move towards zeros for higher gains, such a system would necessarily be a low-gain system.
   - Such a system is a non-minimum phase system and the Bode plot approach cannot be used.
3. Poles of G_pH are outside the unit circle or in the unstable zone:
   - Here for low gains the closed-loop poles would be close to the open-loop poles and would therefore be unstable. Place controller zeros (along with poles) inside the unit circle and pull the loci into the stable zone.

- Since the closed-loop poles are needed to be close to the open-loop zeros, such a system would necessarily be a high-gain system.

4. G_pH has an integrator (pole at $z = 1$):

- To take the steady-state error to 0, an additional integrator is not needed. Hence the controller for this will be like a PD controller.

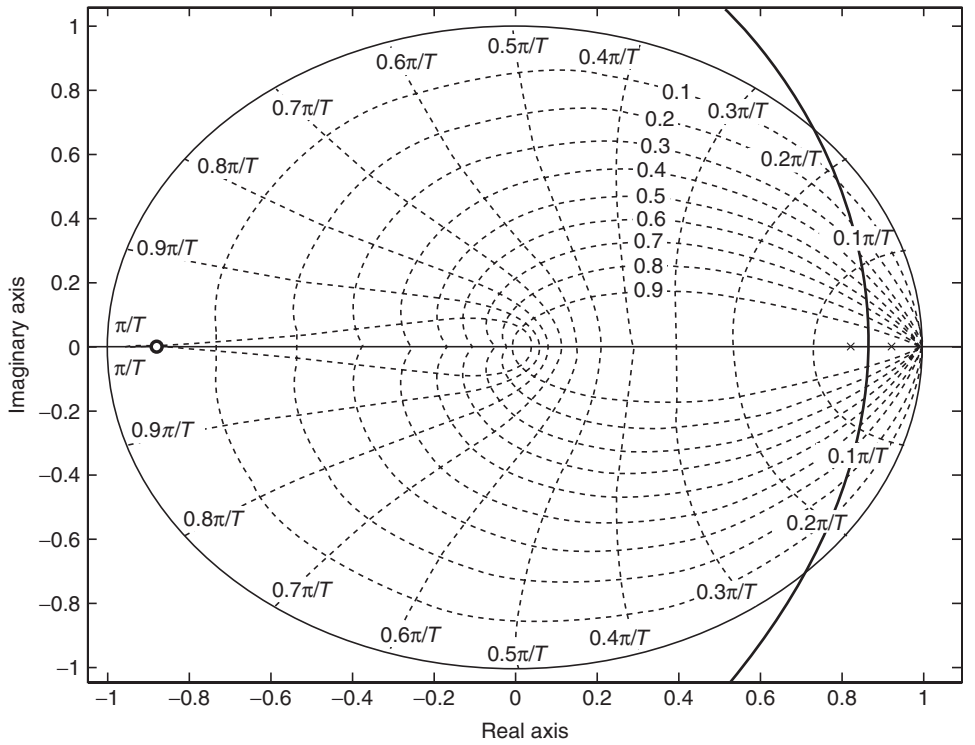


Figure 10.25 | Root locus plot of the open-loop uncompensated system.

EXAMPLE 10.2

The example of the speed control of the DC motor that is discussed in Section 10.2 is considered here wherein a controller based on the root locus approach is designed. The transfer function of the uncompensated DC motor is given as

$$\frac{\omega(z)}{V(z)} = \frac{0.0018233\,(z + 0.9109)}{(z - 0.92)\,(z - 0.8215)}$$

Here the output is the angular speed of the DC motor shaft and the input is the armature voltage. The armature voltage is the control input that is varied to obtain the desired angular speed. The root locus plot of the open-loop uncompensated system is shown in Figure 10.25.

It consists of two poles and a zero within the unit circle. The root loci go out of the unit circle as shown. One can start with the controller that was designed with the Bode approach. The controller designed with the Bode approach is

$$G_c(z) = \frac{6(z-0.9)}{z-1}$$

Using the controller structure without the gain, one obtains

$$G_c(z) = \frac{z-0.9}{z-1}$$

This controller structure is combined with the DC motor plant to obtain the loop transfer function which is given as

$$G_{\text{loop}}(z) = KG_cG_pH = K\frac{z-0.9}{z-1}\left(\frac{0.0018233\cdot(z+0.9109)}{(z-0.92)\cdot(z-0.8215)}\right)$$

The root locus of the above loop transfer function with K as the gain parameter is shown in Figure 10.26.

For a gain of $K = 3.3$, the step response of the closed-loop system is shown in Figure 10.27. Let the controller structure be changed to a PID controller from a PI controller structure. Introduction of the derivative portion will significantly improve the transient response of the system. Introduce a zero at $z = -0.95$ that performs the

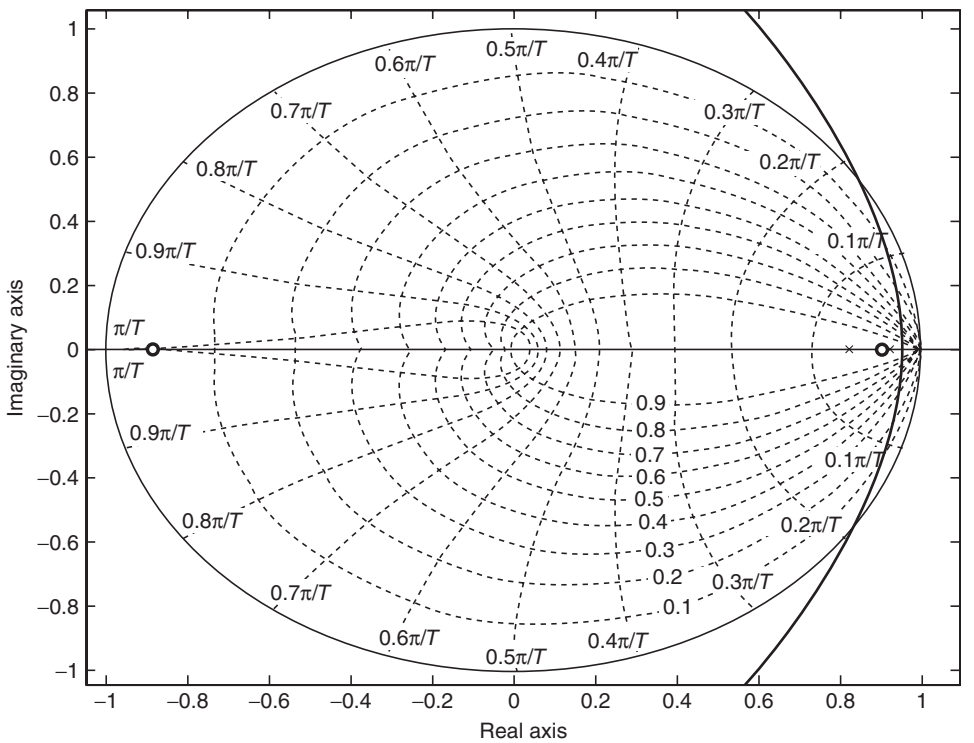


Figure 10.26 | Root locus of the plant with PI controller.

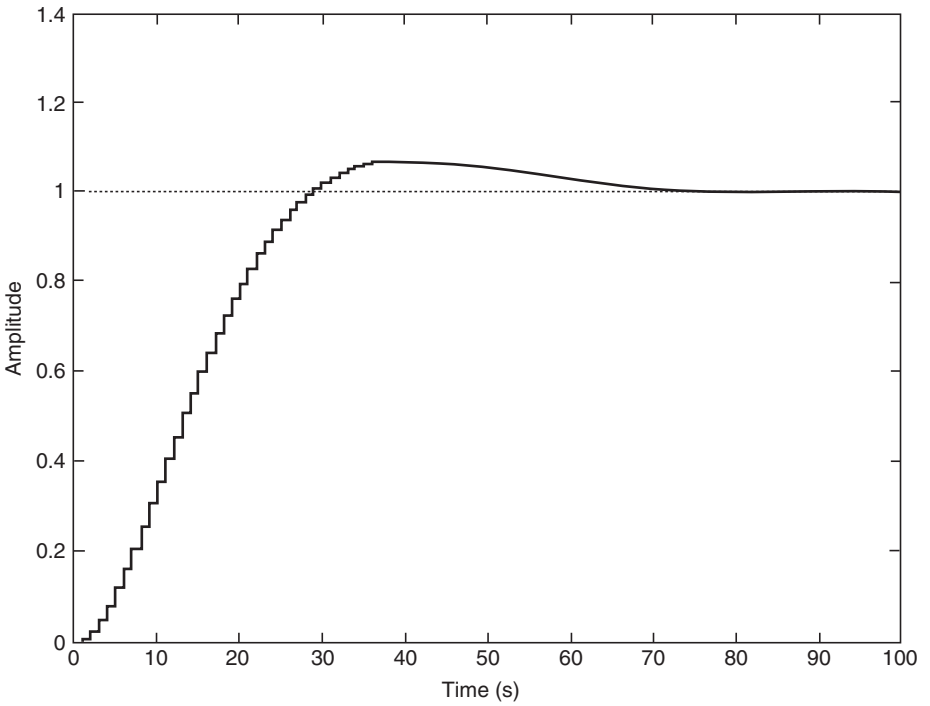


Figure 10.27 | Step response of the closed-loop system for a gain of $K = 3.3$.

derivative action. One cannot introduce only a zero as it will become a non-causal system. Therefore, a pole has to be introduced. Introduce a pole at $z = 0$. The PID controller structure is

$$G_c(z) = \left(\frac{z-0.9}{z-1}\right)\left(\frac{z-0.95}{z}\right)$$

This controller structure is combined with the DC motor plant to obtain the loop transfer function which is given as

$$G_{loop}(z) = KG_cG_pH = K\left(\frac{z-0.9}{z-1}\right)\left(\frac{z-0.95}{z}\right)\left(\frac{0.0018233(z+0.9109)}{(z-0.92)(z-0.8215)}\right)$$

The root locus of the above loop transfer function with gain K as the parameter is shown in Figure 10.28.

The derivative zero is judiciously placed so that the root loci are brought well within the unit circle. The gain can now be chosen higher to improve the transient response. For a gain $K = 227$, the step response of the closed-loop system is shown in Figure 10.29. From the step response as shown in Figure 10.29, it can be observed that the initial transient is due to the derivative part. The steady-state portion of the response is due to the integral part that ensures zero steady-state error. The proportional part acts as the bridge between the derivative part and the integral part. The three operational zones of the PID controller can be easily distinguished from Figure 10.29.

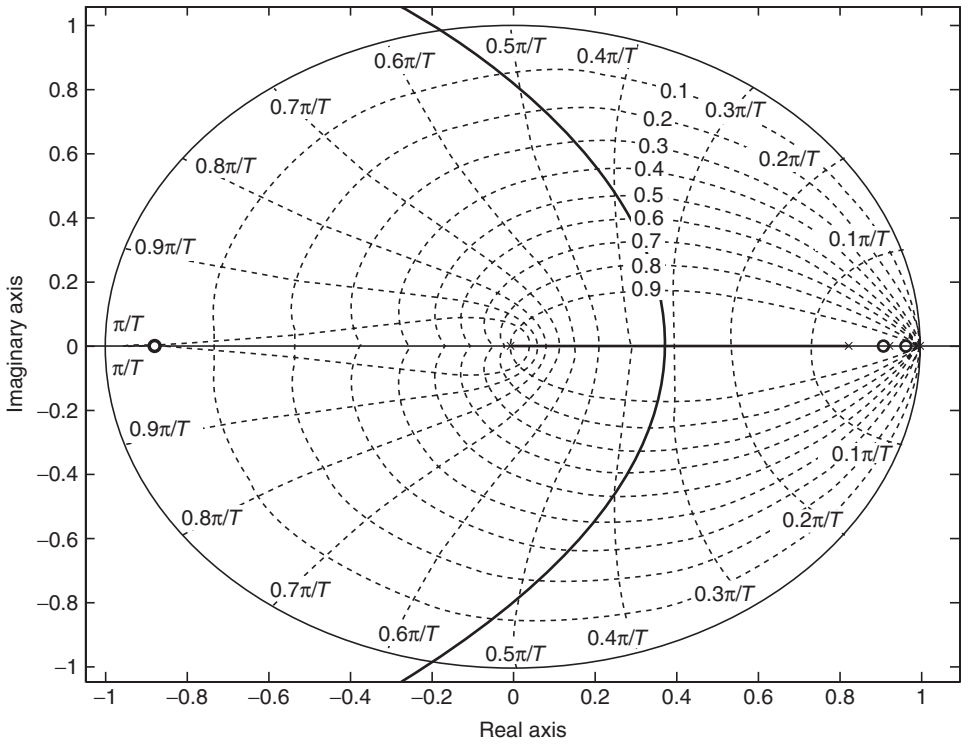


Figure 10.28 | Root locus of the plant with PID controller.

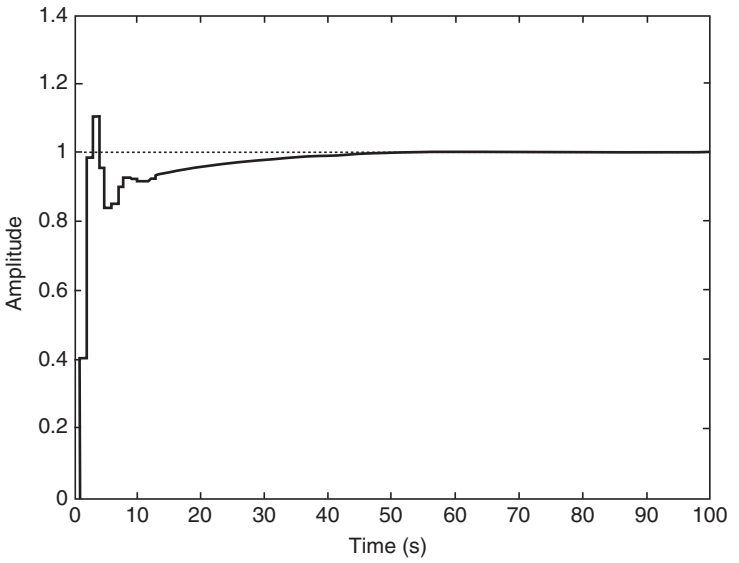


Figure 10.29 | Step response of the closed-loop system with PID controller structure for a gain $K = 227$.

10.5 State Space Method

The state space model is a more general representation of a system compared to the transfer function model. The transfer function assumes zero initial conditions and is applicable only for SISO systems. The state space approach suffers from neither of these limitations. It is applicable for MIMO systems and can even be used for non-linear and time-varying systems. In the discussions to follow, the discrete domain state space controller design methods will be used unless otherwise stated, even though the concepts and principles are equally valid for continuous domain controller design.

The state space equations for a general discrete domain system as discussed in Chapter 9 are in the form

$$x_{k+1} = Ax_k + Bu_k \tag{10.29a}$$

$$y_k = Cx_k + Du_k \tag{10.29b}$$

Equation (10.29a) gives the next state, given the present state and the input. Equation (10.29b) gives the output, given the present state and the inputs. Together the two equations are called the state equations for the system where x_k is the state vector at the discrete time instant kT with sampling time T, u_k the input vector at the discrete time instant kT; A the characteristic matrix for the system; B the input matrix; C the output matrix; D the feed-through matrix.

Any system can be represented in the state space form as indicated above. Chapter 8 gives a detailed discussion on modeling and Chapter 9 discusses the conversion of continuous domain models to discrete domain models. Computer programs like SciLAB and MATLAB have functions that will do these continuous-to-discrete domain transformations thereby eliminating the tedious manual conversion processes. These could be used during the design process.

For any physical system, there are two important conditions that must be met: (a) controllability and (b) observability. If these two conditions are not satisfied then one should re-look at the model and include certain non-idealities that may have been overlooked.

Controllability

A system represented by $x_{k+1} = Ax_k + Bu_k$ is controllable if there exists a sequence of inputs $\{u_0, u_1, u_2, \ldots, u_{n-1}\}$ such that it transfers the system from any initial state to any final state x_n with n being finite.

From the state equation of the system,

$$x_1 = Ax_0 + Bu_0$$

$$x_2 = Ax_1 + Bu_1$$

$$\Rightarrow x_2 = A^2x_0 + ABu_0 + Bu_0$$

By extending to state at nth time interval, one obtains

$$x_n = A^n x_0 + \begin{bmatrix} B & AB & A^2B & \ldots & A^{n-1}B \end{bmatrix} \begin{bmatrix} u_{n-1} \\ u_{n-2} \\ u_{n-3} \\ \vdots \\ u_0 \end{bmatrix}$$

$$(x_n - A^n x_0) = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix} \begin{bmatrix} u_{n-1} \\ u_{n-2} \\ u_{n-3} \\ \vdots \\ u_0 \end{bmatrix}$$

This can be written as

$$\begin{bmatrix} u_{n-1} \\ u_{n-2} \\ u_{n-3} \\ \vdots \\ u_0 \end{bmatrix} = \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}^{-1} (x_n - A^n x_0)$$

It can be seen from the above expression that to transfer the system from any state x_0 to another state x_n, it is possible to find a sequence of inputs to achieve the task only if the matrix

$$\begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix}$$

is invertible. This matrix is called the controllability matrix. This implies that any system can be controlled only if the controllability matrix is invertible.

For any real system modeled correctly, the matrix is always invertible and the system is therefore controllable. If the inverse of the controllability matrix does not exist, it implies that the model of the system is improper.

Observability

A system represented by $x_{k+1} = Ax_k + Bu_k$ as the state equation and $y_k = Cx_k$ as the output equation wherein there is no direct feed-through to the output from the input is said to be observable provided that the initial state x_0 can be calculated from n measurements $y_0, y_1, \dots, y_{n-1}$ with n being finite.

This is actually the dual of controllability. Let the system be transferred to any state x_n by application of a certain sequence of inputs. Then the output at nth time interval is

$$\begin{aligned} y_n &= Cx_n \\ y_{n-1} &= Cx_{n-1} \\ &\vdots \\ y_0 &= Cx_0 \end{aligned}$$

The above equations can be represented as

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^n \end{bmatrix} x_0$$

The initial state x_0 can be obtained as

$$x_0 = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^n \end{bmatrix}^{-1} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The matrix

$$\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^n \end{bmatrix}^{-1}$$

is called the observability matrix. Only if the inverse of the observability matrix exists, can any initial state be computed from n measurements. If the inverse of the observability matrix does not exist, then the system is not observable which again implies incorrect system model.

10.6 Full-State Feedback

In the full-state feedback control, all the state variables of the system are available and therefore fed back to the controller. As the states contain all the information about the plant, the controller does not need additional poles and zeros to include more degrees of freedom. Therefore, the controller needs to be just simple gains for each input. This implies that in the case of the full-state feedback design the order of the system remains unchanged as the controller is only a gain matrix.

In the case of the classical root locus method discussed in the previous section, the closed-loop poles were placed on the root loci portion that is within the unit circle. However, in the case of the full-state feedback control, due to the availability of full-state information, there is increased flexibility in the controller design in that the closed-loop poles of the system can be placed *anywhere within the unit circle*. As a consequence this is a very powerful design tool.

Figure 10.30 depicts the block diagram of a system represented in the state space form with full-state feedback. The plant is represented in the state space form as given by Eqs. (10.29a) and (10.29b). The block schematic representation of these equations is as shown within the marked square in Figure 10.30. The block interconnections are shown as double-lined arrows to indicate that the variables are vectors. The present state vector x_k is fed back and compared with the reference vector r. The difference between the reference vector and the fed back state vector results in the error vector e_k. This error vector is given as the input to the controller K which is a gain matrix. The output of the controller becomes the control input to the plant.

The state space controller is generally classified by the variable that is fed back to the controller and the reference variable. The classification is called the control law. The control law for the above system is

$$u_k = K(r - x_k) \tag{10.30a}$$

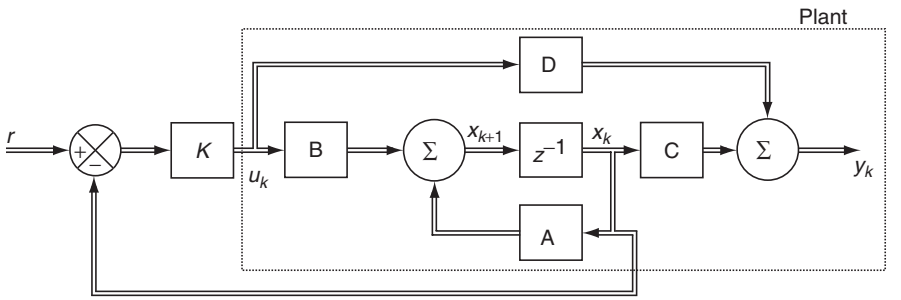


Figure 10.30 | Block diagram describing the full-state feedback.

Further, there are two classes of the controller based on the nature of the reference waveform. They are

1. regulator;
2. tracking.

Regulator

In this case the output needs to be maintained at a fixed set point. The reference in this case is a constant R. From Eq. (10.30a), the control law for the regulator is

$$u_k = K(R - x_k) \tag{10.30b}$$

Substituting the control law in the system $x_{k+1} = Ax_k + Bu_k$, one obtains

$$x_{k+1} = (A - BK)x_k + BKR \tag{10.30c}$$

The system dynamics is decided by $(A - BK)$ where K is the controller. The controller should perform for any value of the constant reference R. Therefore, without loss of generality, the controller designed for $R = 0$ is valid for controller designed for any R. Thus for the regulator design, the control law is simplified by considering $R = 0$. Thus the control law for regulator design is given as

$$u_k = -Kx_k \tag{10.30d}$$

The block schematic of the regulator is shown in Figure 10.31. Figure 10.31(a) shows the regulator structure and Figure 10.31(b) shows the control structure used for controller K design considering $R = 0$.

Tracker

This is a more versatile control structure as compared to the regulator control structure. In this the reference is not fixed but changes with time. This corresponds to large-signal changes in the output. Equation (10.30a)

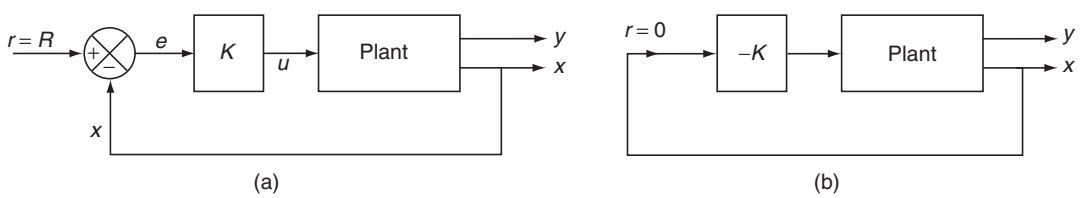


Figure 10.31 | (a) Regulator control structure; (b) regulator structure for controller design with $R = 0$.

is itself the control law where the input and state vectors represent large-signal variations. The controller has to be designed using the large-signal model of the system. Figure 10.32 illustrates the tracker control structure. The large-signal state variable is fed back and compared with the reference r. As r will have the vector dimensions of the controlled outputs y, the block matrix N_x is used to make r compatible with the vector dimensions of the state x. To improve the dynamics, a feed-forward control can also be implemented as shown by directly feeding a part of r to the control input u through the block matrix N_u.

The tracker problem can be considered as the regulator problem plus a feed-forward portion. The controller is designed to take care of the variations of the state variables in the neighborhood of the operating point using the small-signal model of the system. The control input to the plant is the sum of the constant reference plus the correction action to reject the perturbations in the neighborhood of the operating point. Thus,

$$u = U + \underset{\sim}{u} = R + \underset{\sim}{u}$$

where R is the constant reference indicating nominal operating point and $\underset{\sim}{u}$ is the small-signal variations of the control input about the nominal control input $U = R$. The feedback state variable $\underset{\sim}{x}_k$ represents the small-signal variations about the operating point and the small-signal reference is zero. From Eq. (10.30d), the control law for the regulator problem is

$$\underset{\sim}{u}_k = -K\underset{\sim}{x}_k$$

where $\underset{\sim}{x}_k = x_k - X$ with x_k being the actual large-signal state value and X the state value at the nominal operating point. Figure 10.32 illustrates the tracker control structure by splitting the tracker problem of Figure 10.32 into the regulator problem and the feed-forward portion.

Figure 10.33(a) shows the block schematic of the regulator plus feed-forward control. The controller K corrects for the state perturbation around the operating point. In the steady state, $\underset{\sim}{x}_k$ will become zero. X is the steady-state value or the nominal operating point corresponding to the input $U = R$. The steady-state value X is obtained from the steady-state model of the plant with $U = R$ as the input. Figure 10.33(b) shows a simplified regulator plus feed-forward control structure.

To design the controller K, the small-signal model of the plant is sufficient as the controller handles only the small-signal dynamics. The control block diagram used for designing the controller K is shown in Figure 10.34 wherein the plant is only the small-signal model.

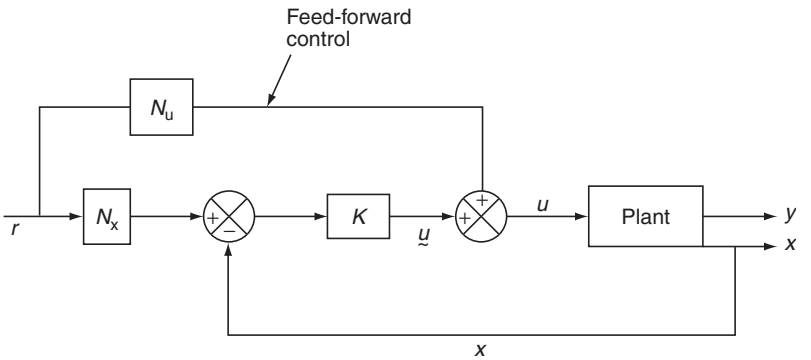


Figure 10.32 | Block diagram of the tracker control structure.

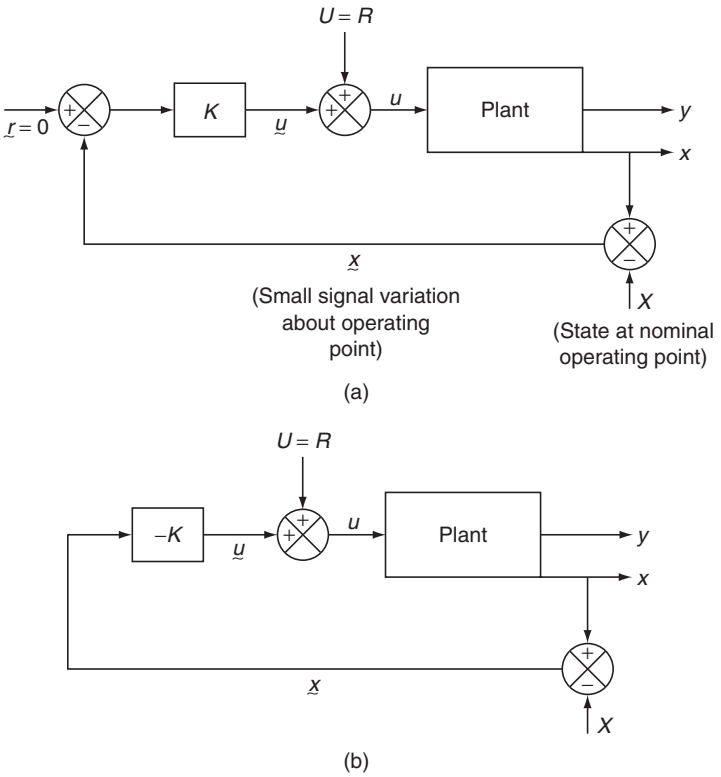


Figure 10.33 (a) Block diagram of regulator plus feed-forward control structure; (b) reduced regulator plus feed-forward control structure.

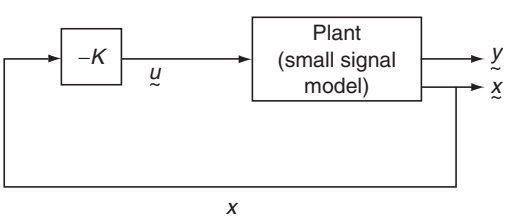


Figure 10.34 Block diagram for controller design for tracker problem.

10.7 Regulator Design by Pole Placement

This section discusses the controller design for the regulator control problem. The control law is given in Eq. (10.31) as

$$u_k = -Kx_k \tag{10.31}$$

The plant model is given as

$$x_{k+1} = Ax_k + Bu_k \tag{10.32a}$$

$$y_k = Cx_k + Du_k \tag{10.32b}$$

Equation (10.32a) gives the dynamics of the system. The input to this system is as given in Eq. (10.31). Incorporating the control law into the state equation [Eq. (10.32a)], one obtains

$$x_{k+1} = Ax_k - BKx_k = (A - BK)x_k \tag{10.33}$$

The closed-loop dynamics of the system is determined by $A - BK$. From Eq. (10.33), the characteristic equation is obtained as the determinant of $zI - (A - BK)$. Thus,

$$|zI - (A - BK)| = 0 \tag{10.34}$$

This gives the characteristic equation of the full-state feedback system in the following equivalent polynomial form as

$$\xi_p(z) = z^n + \beta_{n-1}z^{n-1} + \cdots + \beta_1 z + \beta_0 = 0 \tag{10.35}$$

The coefficients β_i will be functions of K and the parameters of the system. The roots of Eq. (10.35) give the closed-loop pole location of the full-state feedback system.

From the model of the plant, the order of the system is known and therefore the number of closed-loop poles to be selected is also known. Let the closed-loop poles for the full-state feedback system be selected as $\lambda_1, \lambda_2, \ldots, \lambda_n$ within the unit circle. From the selected or desired closed-loop pole location values, the desired characteristic equation is

$$(z - \lambda_1)(z - \lambda_2)(z - \lambda_3)... = 0 \tag{10.36}$$

which can be represented in the equivalent polynomial form as

$$\xi_d(z) = z^n + \alpha_{n-1}z^{n-1} + \cdots + \alpha_1 z + \alpha_0 = 0 \tag{10.37}$$

where α_i are functions of the desired closed-loop pole locations.

Equation (10.35) gives the characteristic equation that relates with the system parameters and the controller gain K. Equation (10.37) gives the characteristic equation that relates with the chosen or desired closed-loop pole locations for the full-state feedback system. Equating these two characteristic equations and comparing the coefficients, a set of n simultaneous linear equations is obtained with K as the unknown (K is the gain matrix containing n gains). The solution to this set of simultaneous equations gives the values of the controller gain matrix K which will result in the selected or chosen closed-loop pole locations, $\lambda_1, \lambda_2, \ldots, \lambda_n$. Computer programs can be used to obtain the solution for the set of n simultaneous equations. One of the algorithms is based on the Ackerman's formula which is given as

$$K = \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \begin{bmatrix} B & AB & A^2B & \ldots & A^{n-1}B \end{bmatrix}^{-1} \cdot \xi_d(A) \tag{10.38}$$

where $\xi_d(A) = A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0$ is obtained from the desired characteristic equation of $\xi_d(z)$ by replacing the variable z by A. The design is algorithmic in nature and summarized in the following steps:

Step 1: Model the plant in the continuous domain.

Step 2: Convert the model from the continuous-to-discrete domain as discussed in Chapter 9 and obtain the A, B, C and D matrices.

Step 3: Select all the desired closed-loop poles locations within the unit circle of the z-plane. The closer the poles are to the origin of the unit circle, faster is the speed of response. In the limit, if all the closed-loop poles are placed at the origin, then the system becomes an nth-order delay. This is the fastest response that one can obtain from the system. Such a controller where all the $\lambda_i = 0$ is called a dead beat controller. It should, however, be noted that closer the closed-loop poles to the origin, larger will be the resulting controller gains. This implies high control energy requirement which the plant should be capable of providing.

Step 4: Obtain the controller gains K by providing A, B and λ as inputs to the Ackerman's formula or an equivalent algorithm for solving the n simultaneous equations.

Step 5: Obtain the step response of the closed-loop system. Use a computer program to plot the step response.

Step 6: Iterate Steps (3)–(5) till the step response satisfactorily meets all the performance requirements.

EXAMPLE 10.3

Consider the example of the DC motor speed control, the model of which is described in Section 10.2. The controller for the discrete domain state space model given in the above-mentioned example is designed here. For the parameters of the DC motor given in the earlier section, the state space model is given as

$$A = \begin{bmatrix} 0.92294 & -0.00348 \\ 0.08698 & 0.81856 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.03844 & 0.00182 \\ 0.00182 & -0.90629 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$x = \begin{bmatrix} ia; \omega \end{bmatrix} \quad u = \begin{bmatrix} v; TL \end{bmatrix}$$

where x is the state vector and u is the input vector.

The plant has two inputs: (a) v – the armature voltage that actually controls the speed, (b) T_L – the load torque. The load torque acts as a disturbance input to the system. The control input is only the armature voltage. Hence the control input vector is [v] considering the disturbance input $T_L = 0$ for purposes of controller synthesis without loss of generality. As it is a second-order system, let the desired closed-loop poles be chosen as $z = 0$ and $z = 0$. This means that both the closed-loop poles are placed at the origin of the unit circle. This will result in dead beat control and will also give the fastest response. As the order of the system is two, in two times the sample time, the system will reach steady state. The gains are computed by means of a computer program which solves the Ackerman's formula. If the function computing the Ackerman's formula is called "acker", then

$$K = \text{acker}(A, B, P)$$

where P is the vector of desired closed-loop pole locations and in this case $P = [0, 0]$. The state and the input responses of the closed-loop system are given in Figure 10.35. The value of the controller gain vector is

$$K = [35.22729, 212.4682]$$

From Figure 10.35, it can be observed that the steady state is reached in two sample times. As the reference input in the regulator design is assumed to be zero, the states will reach the zero reference value in two sample times. Figure 10.36 shows the state and input response for an arbitrary location of the closed-loop poles within the unit circle. For $z = 0.1$ and $z = 0.3$ as the desired closed-loop pole locations,

$$K = [29.29581, 118.1327]$$

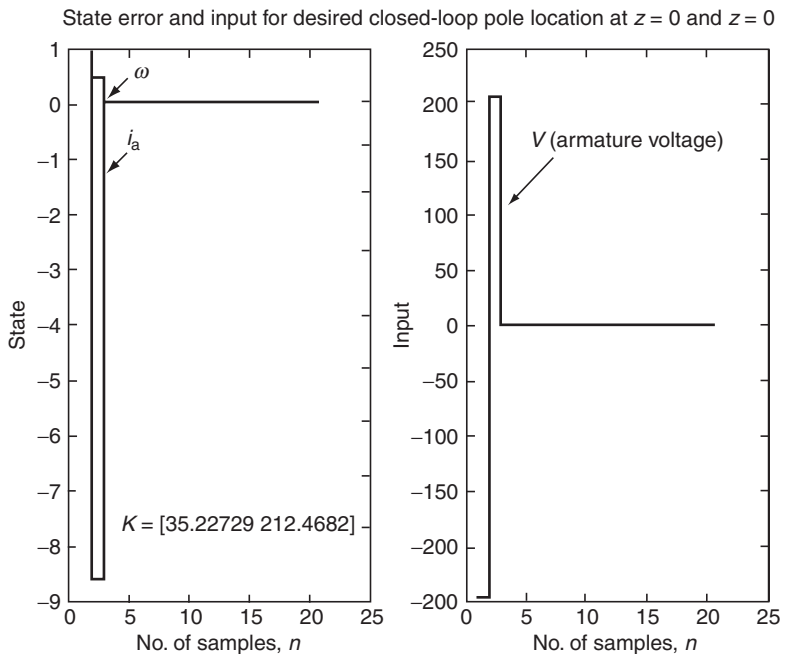

State error and input for desired closed-loop pole location at z = 0 and z = 0
ω
i_a
K = [35.22729 212.4682]
State
No. of samples, n
V (armature voltage)
Input
No. of samples, n

Figure 10.35 | State error and input for dead beat controller.

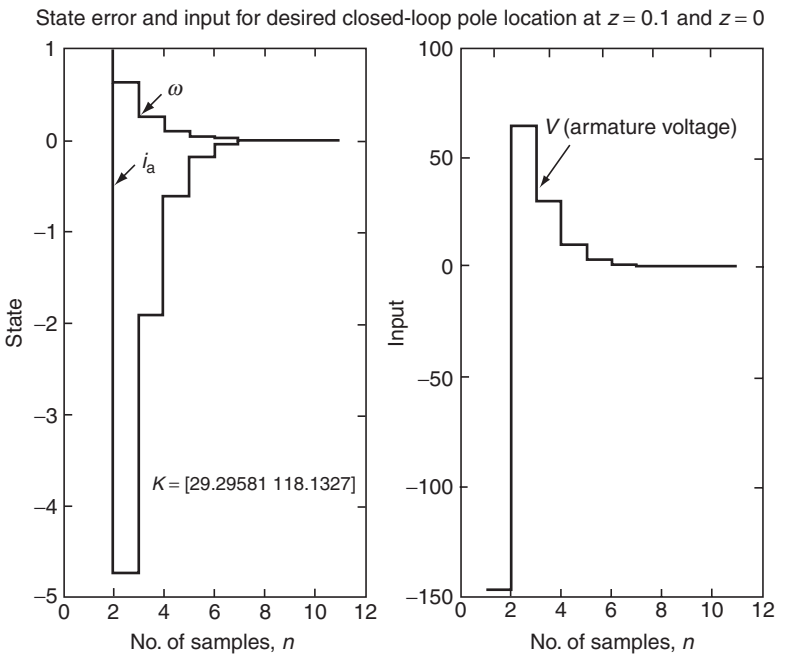

State error and input for desired closed-loop pole location at z = 0.1 and z = 0
ω
i_a
K = [29.29581 118.1327]
State
No. of samples, n
V (armature voltage)
Input
No. of samples, n

Figure 10.36 | State error and input for arbitrary closed-loop pole placement.

10.8 Estimator Design

In full-state feedback, it has been considered that all the state variables of a system are measurable. In many practical situations all the state variables may not be easily measurable or may not be accessible due to physical limitations. In such cases, the states of the system will have to be estimated based on the information available in the measurable states. Such a unit is called an *estimator* or an *observer*. The estimator or observer that observes the state values is in actuality a replacement for the physical sensors that would generally be used in the case of measurable states.

A model of the plant, which should be proper, is run alongside the plant. Control inputs are given to this model as well as the plant. The controller uses the estimated state variable values to compute the control action. The output of the plant model is compared with the actual plant output and the error is passed through a gain vector G called the estimator gain. The output of the estimator is used to apply corrective action to the estimator states.

Figure 10.37 depicts the full-state feedback system with an estimator. Only few of the states may be measurable. The measured states are brought out as the system output y_k as indicated. The estimator gives an estimate of the measured output $\hat{y}_k$. The difference between the measured output and the estimated output is passed through an estimator gain G and given as a local corrective negative feedback to the estimator as shown. The estimator uses the plants model A, B, C and D and the actual input to the plant u to estimate the states $\hat{x}_k$. The estimated $\hat{x}_k$ is fed back to the controller K to perform the control action. Here G is a simple gain vector. The variables with the hat "^" represent the estimated values. From Figure 10.37, it can be observed that the state equation for the estimator is

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + G(y_k - \hat{y}_k) \tag{10.39}$$

On re-arranging, one obtains

$$\hat{x}_{k+1} = (A - GC)\hat{x}_k + Bu_k + Gy_k \tag{10.40}$$

Equation (10.40) is the estimate equation. It is called the *prediction estimator*. $\hat{x}_k$ represents the estimator state and u_k and y_k form the inputs to the estimator. The estimator dynamics is determined by $A - GC$.

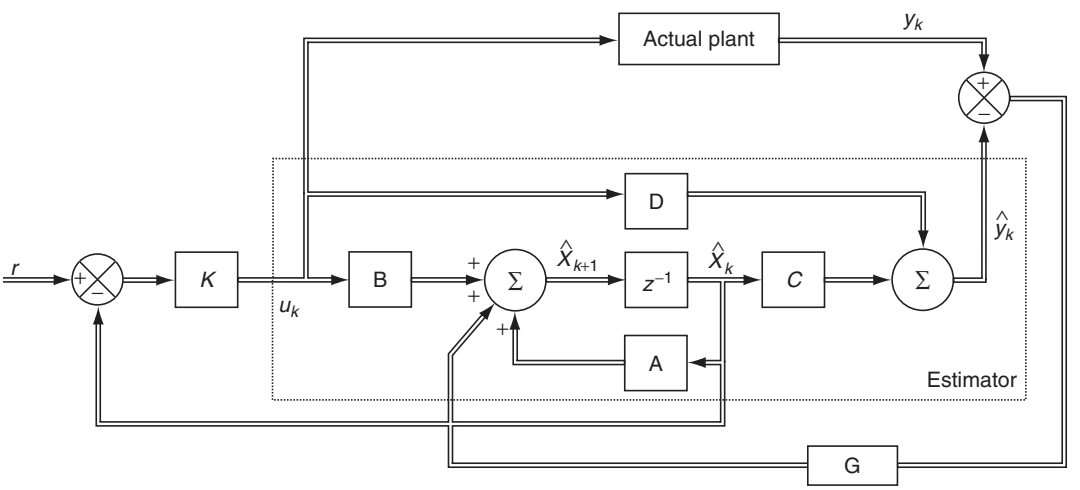


Figure 10.37 The schematic of the full-state feedback system with an estimator.

Estimator Error Dynamics

The estimator error gives a measure of the closeness of the actual state and the estimated state. Defining the estimator error as the difference between the actual state variable and the estimated state variable, one obtains

$$e_k = x_k - \hat{x}_k \tag{10.41}$$

$$\begin{aligned} e_{k+1} &= x_{k+1} - \hat{x}_{k+1} \\ &= Ax_k + Bu_k - (A - GC)\hat{x}_k - Bu_k - GCx_k \\ &= (A - GC)e_k \end{aligned} \tag{10.42}$$

Equation (10.42) gives the error dynamics of the system which is same as the estimator dynamics. It is desired that the error should not increase with time. Hence if the estimator is stable, the error will also be stable and decay to zero. The equation is of a similar form as $A - BK$ which gives the closed-loop dynamics of the full-state feedback system. The estimator may be viewed as a local closed-loop system with dynamics determined by $A - GC$. The method described to design the controller K for the full-state feedback can be used to find the value of G.

The controller K is designed using the Ackerman's formula or a similar algorithm. In the controller design, the closed-loop dynamics is determined by $A - BK$ where A is the $n \times n$ matrix for an n-state system, B the $n \times m$ matrix for m control inputs, K the $m \times n$ matrix. In the case of the estimator, the estimator dynamics is decided by $A - GC$ where A is the $n \times n$ matrix, C the $m \times n$ matrix (in a practical system there should be one control input for every controlled output; thus for m control inputs, the maximum number of controlled outputs is m), G the $n \times m$ matrix.

To dimensionally match with the components of $A - BK$,

$$(A - GC)^{\mathrm{T}} = A^{\mathrm{T}} - C^{\mathrm{T}}G^{\mathrm{T}} \tag{10.43}$$

By transposing $A - GC$, the dynamics of the estimator is not affected, but it is dimensionally transformed exactly to the form of $A - BK$. Now the same Steps (3)–(6) given in Section 10.7 can be applied to design the estimator gain G^{T} by using A^{T}, C^{T} and the desired estimator closed-loop pole locations. The design steps are:

Step 1: Select all the desired closed-loop poles locations λ_{e} of the estimator within the unit circle of the z-plane. As a general guideline, the estimator should be faster than the plant. So for obtaining G, the closed-loop poles for the estimator will be chosen closer to the origin compared to the closed-loop poles chosen for the controller K design.

Step 2: Obtain the estimator gains G by providing A^{T}, C^{T} and λ_{e} as inputs to the Ackerman's formula as given in Eq. (10.38) or an equivalent algorithm for obtaining G^{T}. In Ackerman's formula A is replaced by A^{T}, B is replaced by C^{T} and K is replaced by G^{T}.

Step 3: Obtain the error step response of the estimator system. Use an appropriate computer program to plot the step response.

Step 4: Iterate Steps (1)–(3) till the step response satisfactorily meets all the performance requirements.

EXAMPLE 10.4

For the DC motor plant, the prediction estimator gain G is designed by first placing the estimator poles. The fastest estimation is done when all closed-loop estimator poles are at the origin. In general, the estimator poles should be placed closer to the origin than the controller closed-loop poles. Figure 10.38 shows the error between the actual and the estimated states for estimator poles placed at the origin and at 0.2. Here again the Ackerman's formula is used. The function is called with the following parameters to the Ackerman's function:

$$G^{\mathrm{T}} = \mathrm{acker}(A^{\mathrm{T}}, C^{\mathrm{T}}, P_{\mathrm{e}})$$

where P_{e} is the vector of desired closed-loop estimator pole locations.

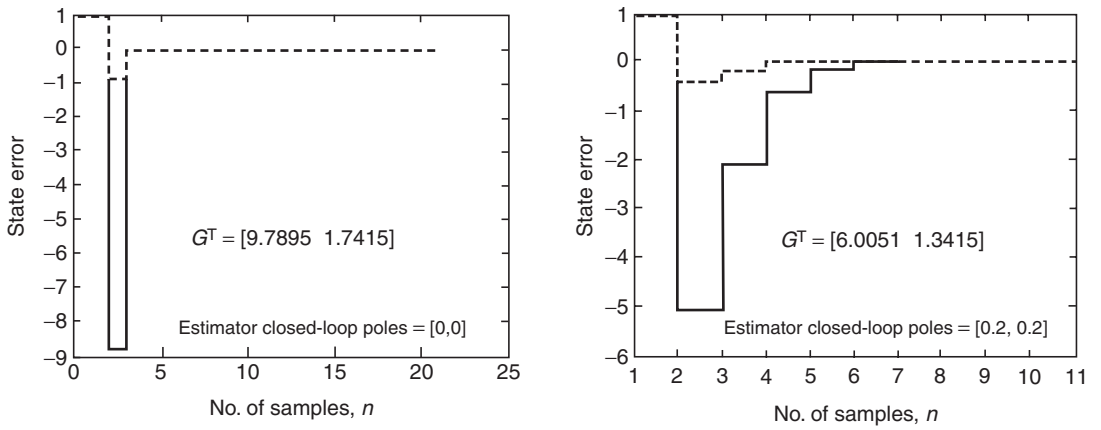


Figure 10.38 State errors for two different estimator closed-loop pole locations.

Separation Principle

In the full-state feedback system, there are two sets of dynamics: (a) closed-loop plant dynamics as decided by $A - BK$ and (b) the estimator dynamics as decided by $A - GC$. The question that needs to be answered is whether there is any sequence in which the estimator and controller gains should be designed? For example, should the estimator gain be designed first and then the controller gain or the other way round? Or can both be designed separately and independent of each other for a given system?

For a system that is represented as

$$x_{k+1} = Ax_k + Bu_k$$

applying the regulator control law $u_k = -Kx_k$, the closed-loop system with full-state feedback is given as

$$x_{k+1} = Ax_k - BKx_k = (A - BK)x_k$$

Instead of actual state feedback, if the estimated states are fed back then

$$x_{k+1} = Ax_k - BK\hat{x}_k \tag{10.44}$$

Due to the use of the estimated states rather than the actual states for feedback, the state error is $e_k = x_k - \hat{x}_k$ and therefore, $\hat{x}_k = x_k - e_k$. Substituting this in Eq. (10.44), the system state equation is given as

$$x_{k-1} = Ax_k - BKx_k + BKe_k \tag{10.45}$$

where the error becomes an implicit input to the system. The error dynamics is given by Eq. (10.42). Considering x_k and e_k as the composite state of the system, the overall system dynamics with the estimator is given by the following state equation which is a combination of Eqs. (10.42) and (10.44).

$$\begin{bmatrix} x_{k+1} \\ e_{k+1} \end{bmatrix} = \begin{bmatrix} A - BK & BK \\ 0 & A - GC \end{bmatrix} \begin{bmatrix} x_k \\ e_k \end{bmatrix} \tag{10.46}$$

From the state equation [Eq. (10.46)], the characteristic equation may be written as

$$|zI - A + BK||zI - A + GC| = 0 \tag{10.47}$$

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the closed-loop pole locations of the controller and $\lambda_{e1}, \lambda_{e2}, \ldots, \lambda_{en}$ are the estimator pole locations, then from Eq. (10.47), the characteristic equation can be represented as

$$\underbrace{\{(z-\lambda_1)(z-\lambda_2)\cdots\cdots(z-\lambda_n)\}}_{\text{Closed-loop poles}}\underbrace{\{(z-\lambda_{e1})(z-\lambda_{e2})\cdots\cdots(z-\lambda_{en})\}}_{\text{Estimator poles}} = 0 \tag{10.48}$$

Equation (10.48) is the characteristic equation of the complete full-state feedback system with estimator and it can be split into the following two equations:

$$\underbrace{\{(z-\lambda_1)(z-\lambda_2)\cdots\cdots(z-\lambda_n)\}}_{\text{Closed-loop poles}} = 0$$

$$\underbrace{\{(z-\lambda_{e1})(z-\lambda_{e2})\cdots\cdots(z-\lambda_{en})\}}_{\text{Estimator poles}} = 0$$

It can be observed that the choice of the closed-loop poles is not dependent on the choice of the estimator poles and vice-versa. This implies that the controller design K and the estimator gain design G can be done separately and independent of each other. This is called the "Separation Principle".

Current Estimator

The state equation for the prediction estimator is given in Eq. (10.40) and is reproduced here:

$$\hat{x}_{k+1} = (A - GC)\hat{x}_k + Bu_k + Gy_k$$

In the prediction estimator, the values of the output y_k (i.e., the measured states) correspond to the previous time interval to estimate the next state estimate. In other words, the measured states of the previous time interval are used to predict the next state values. Hence, it is called "prediction estimator". In the case of the current estimator, the latest or the measured states of the current time interval are used to correct the estimate in the current time interval. Therefore, the name "current estimator". The current estimator is a two-step process. The two steps are:

1. time update;
2. measurement update.

During the time update process, the next state of the state vector is estimated, that is, the $(k + 1)$th state is estimated using the kth values of the state and measured outputs. This process of estimate update from k to $k + 1$ is called the *time update*. Next, the fresh values of the outputs in the current time interval [i.e., $(k + 1)$]

is used to correct the predicted estimate to obtain the current estimate. This process is called the *measurement update*. Thus the two processes of the current estimation are:

1. **Time Update**:

$$\tilde{x}_{k+1} = A\hat{x}_k + Bu_k \tag{10.49}$$

where $\tilde{x}_{k+1}$ is the time-updated estimate.

2. **Measurement Update**:

$$\hat{x}_{k+1} = \tilde{x}_{k+1} + G(y_{k+1} - C\tilde{x}_{k+1}) \tag{10.50}$$

where the intermediate time-updated state value $\tilde{x}_{k+1}$ is corrected with fresh or current measurement y_{k+1} to obtain the current estimate $\hat{x}_{k+1}$.

Substituting Eq. (10.49) into Eq. (10.50), one obtains

$$\hat{x}_{k+1} = (A - GCA)\hat{x}_k + Gy_{k+1} + (B - GCB)u_k \tag{10.51}$$

This is the consolidated current estimator's estimate equation. The error dynamics for the current estimator is obtained in a similar manner as discussed for the simple prediction estimator. The error is given as

$$e_k = x_k - \hat{x}_k$$

$$e_{k+1} = x_{k+1} - \hat{x}_{k+1}$$

$$e_{k+1} = Ax_k + Bu_k - (A - GCA)\hat{x}_k - Gy_{k+1} - (B - GCB)u_k$$

$$e_{k+1} = (A - GCA)e_k \tag{10.52}$$

Equation (10.52) gives the error dynamics of the system. The design of the current estimator gain is exactly similar to the prediction estimator gain. By transposing $A - GCA$, the dynamics of the estimator is not affected, but it is dimensionally transformed exactly to the form $A - BK$:

$$(A - GCA)^{\mathrm{T}} = A^{\mathrm{T}} - (A^{\mathrm{T}}C^{\mathrm{T}})G^{\mathrm{T}}$$

The same steps of the predication estimator design can be applied to design the estimator gain G^{T} by using A^{T}, $(A^{\mathrm{T}}C^{\mathrm{T}})$ and the desired estimator closed-loop pole locations. The design steps are as follows:

Step 1: Select all the desired closed-loop poles locations λ_{e} of the estimator within the unit circle of the z-plane. The estimator should be faster than the plant. So for obtaining G, the closed-loop poles for the estimator will be chosen closer to the origin compared to the closed-loop poles chosen for the controller K design.

Step 2: Obtain the estimator gains G by providing the A^{T}, C^{T} and λ_{e} as inputs to the Ackerman's formula as given in Eq. (10.38) or an equivalent algorithm for obtaining G^{T}. In the Ackerman's formula A is replaced by A^{T}, B is replaced by $A^{\mathrm{T}}C^{\mathrm{T}}$ and K is replaced by G^{T}.

Step 3: Obtain the error step response of the estimator system. Use an appropriate computer program to plot the step response.

Step 4: Iterate Steps (1)–(3) till the step response satisfactorily meets all the performance requirements.

It should be noted that the current estimator is the basis for many optimal estimators. The principles of the current estimator such as time update and measurement update are used in optimal estimators like the Kalman filter as is discussed in the next chapter.

EXAMPLE 10.5

For the DC motor plant, the current estimator gain G is designed by first placing the estimator poles. Figure 10.39 shows the error between the actual and the estimated states for estimator poles placed at 0.2. Here again the Ackerman's formula is used. The function is called with the following parameters to the Ackerman's function:

$$G^{\mathrm{T}} = \mathrm{acker}(A^{\mathrm{T}}, A^{\mathrm{T}}C^{\mathrm{T}}, P_{\mathrm{e}})$$

where P_{e} is the vector of desired closed-loop estimator pole locations.

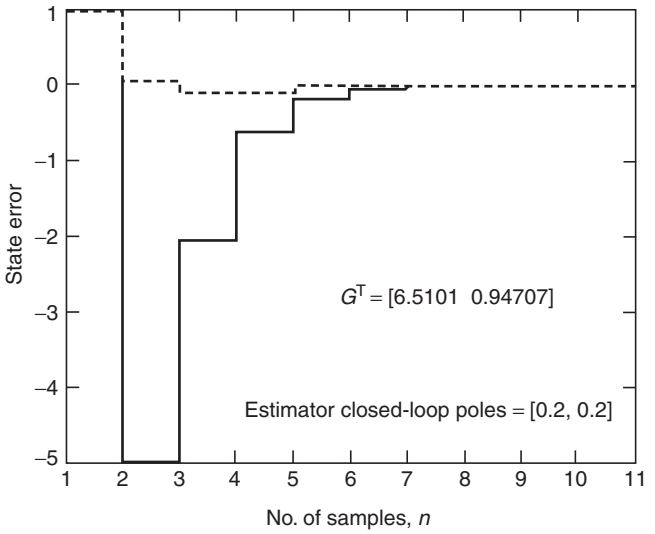


Figure 10.39 | State errors for current estimator.

Reduced Order Estimator

If some of the state variables are available for measurement, it makes sense to estimate only the remaining state variables instead of estimating all the state variables. The state variables that are measured are fed back as it is and the non-measurable state variables are estimated through an estimator and the estimated states along with the measured states are fed back. This reduces the errors because some of the state values are the actual values existing in the plant. Such an estimator that estimates only part of the state vector is called a "reduced order estimator".

Let the state vector be partitioned into two parts: (a) a part x_{m} where the state variables are measurable and (b) a part x_{e} where the state variables need to be estimated. Thus,

$$x = \begin{bmatrix} x_{\mathrm{e}} \\ x_{\mathrm{m}} \end{bmatrix}$$

The system $x_{k+1} = Ax_k + Bu_k$ can now be represented as

$$\begin{bmatrix} x_{\mathrm{e}} \\ x_{\mathrm{m}} \end{bmatrix}_{k+1} = A \begin{bmatrix} x_{\mathrm{e}} \\ x_{\mathrm{m}} \end{bmatrix}_{k} + Bu_k \tag{10.53}$$

where

$$A = \begin{bmatrix} A_{ee} & A_{em} \\ A_{me} & A_{mm} \end{bmatrix}, \quad B = \begin{bmatrix} B_e \\ B_m \end{bmatrix}, \quad x = \begin{bmatrix} x_e \\ x_m \end{bmatrix}$$

Now,

$$\begin{bmatrix} x_e \\ x_m \end{bmatrix}_{k+1} = \begin{bmatrix} A_{ee} & A_{em} \\ A_{me} & A_{mm} \end{bmatrix} \begin{bmatrix} x_e \\ x_m \end{bmatrix}_k + \begin{bmatrix} B_e \\ B_m \end{bmatrix} u_k \tag{10.54}$$

From Eq. (10.54), the state equation for the state variables to be estimated is given as

$$x_{ek+1} = A_{ee} x_{ek} + \underbrace{A_{em} x_{mk} + B_e u_k}_{B_n u_{nk}}$$

$A_{em} x_{mk} + B_e u_k$ can equivalently be represented as $B_n u_{nk}$ because x_{mk} is measurable. Thus

$$B_n u_{nk} = \begin{bmatrix} A_{em} & B_e \end{bmatrix} \begin{bmatrix} x_{mk} \\ u_k \end{bmatrix} \tag{10.55}$$

From Eq. (10.54), the state equation for the measurable state variables is given as

$$x_{mk+1} = A_{mm} x_{mk} + A_{me} x_{ek} + B_m u_k$$

$$x_{mk+1} - A_{mm} x_{mk} - B_m u_k = A_{me} x_{ek}$$

$$\underbrace{x_{mk+1} - A_{mm} x_{mk} - B_m u_k}_{y_{nk}} = \underbrace{A_{me} x_{ek}}_{C_n x_{ek}}$$

The system given by Eq. (10.54) can now be represented only in terms of the state variables that need to be estimated. A full-order state estimator can be built for x_e with u_n as the equivalent input and y_n as the corresponding estimated output. This estimated output will be compared against the actual value of y_n which is completely measurable. As y_n includes a $k + 1$ term, which lies in the future, the rest of the estimator inputs are delayed by one time instant to make this the current time instant term.

The system given by Eq. (10.54) is now represented as

$$x_{ek+1} = A_{ee} x_{ek} + B_n u_{nk}$$

$$y_{nk} = C_n x_{ek} \tag{10.56}$$

where

$$u_{nk} = \begin{bmatrix} x_{mk} \\ u_k \end{bmatrix}; \quad B_n = \begin{bmatrix} A_{em} & B_e \end{bmatrix}; \quad C_n = \begin{bmatrix} A_{me} \end{bmatrix}$$

The full-order prediction or current estimator for the system given by Eq. (10.56) is as explained in the previous sections. The design steps are as follows:

Step 1: Select all the desired closed-loop poles locations λ_e of the reduced order estimator within the unit circle of the *z*-plane. The estimator should be faster than the plant. So for obtaining *G*, the closed-loop poles for the estimator will be chosen closer to the origin compared to the closed-loop poles chosen for the controller *K* design.

Step 2: Obtain the estimator gains G by providing the A_{ee}^{T}, C_n^{T} and λ_e as inputs to the Ackerman's formula as given in Eq. (10.38) or an equivalent algorithm for obtaining G^{T}. In the Ackerman's formula A is replaced by A_{ee}^{T}, B is replaced by C_n^{T} and K is replaced by G^{T}.

Step 3: Obtain the error step response of the estimator system. Use an appropriate computer program to plot the step response.

Step 4: Iterate Steps (1)–(3) till the step response satisfactorily meets all the performance requirements.

EXAMPLE 10.6

For the DC motor plant, let the armature current be estimated and the shaft angular speed be measured. There is only one state to be estimated, the armature current. The reduced order estimator gain G is designed by placing the estimator pole at $z = 0.3$. Figure 10.40 shows the error between the actual and the estimated state. Here again the Ackerman's formula is used. The function is called with the following parameters to the Ackerman's function:

$$G^{T} = \text{acker}(A_{ee}^{T}, A_{ee}^{T} A_{me}^{T}, P_e)$$

where P_e is the vector of desired closed-loop estimator pole location.

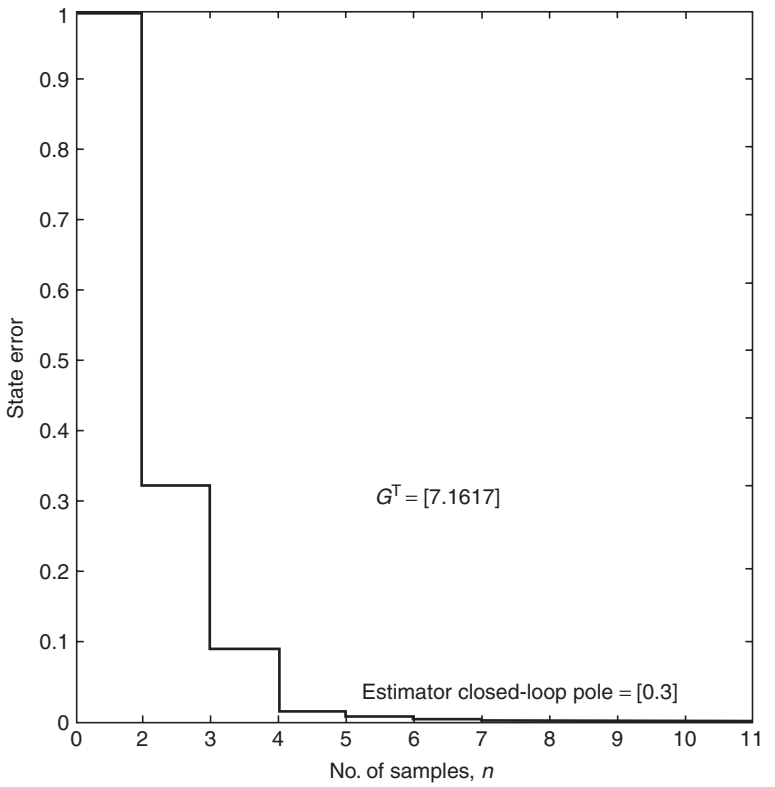


Figure 10.40 | State errors for reduced order estimator.

10.9 Tracker: Controller Design

The controller designed for the regulator problem drives the state values to 0 or a constant reference R. The task of the controller now consists of tracking a varying reference.

Figure 10.41 shows the block diagram of the tracking system. Here r is the reference input which may vary. The length of the reference vector r will be same as the number of controlled outputs. The length of the input vector also will be the same as the number of controlled outputs as each controlled output needs at least one control input. The matrix N_x makes the r vector compatible with the state vector. This part of the input handles the disturbances of the system in the neighborhood of the operating point. The N_u matrix is responsible for setting the operating point so that the controller has a 0 output when the system is at steady state. Thus, the controller is inactive when the system has reached steady state since the required control input to maintain the output at r is provided by the N_u matrix gain. To find the values of N_x and N_u, one may proceed as follows. At steady state,

$$y = r; \quad y = CX; \quad X = N_x^* r$$

because the error e is 0. Therefore

$$CN_x = I \tag{10.57}$$

The control law is given as

$$u_k = KN_x r - Kx_k + N_u r = (KN_x + N_u)r - Kx_k \tag{10.58}$$

Substituting the control law given above into the system equation $x_{k+1} = Ax_k + Bu_k$, one obtains

$$x_{k+1} = (A - BK)x_k + B(KN_x + N_u)r$$

The feedback from the state elements is not affected by the addition of these two matrices and hence the system dynamics remain the same and are dictated by $A - BK$.

Since in steady state, $e = 0$, $u_k = N_u \times r$

$$x_{\infty+1} = x_\infty = Ax_\infty + Bu_\infty$$

$$(A - I)x_\infty + Bu_\infty = \Theta$$

Substituting $x_\infty = N_x r$ and $u_\infty = N_u r$, one obtains

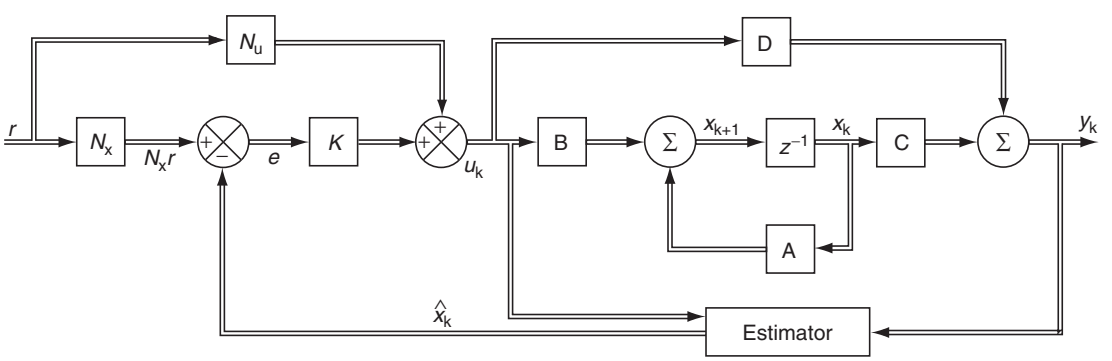


Figure 10.41 | Block schematic of the tracking control system.

$$(A - I)N_x + BN_u = \Theta \tag{10.59}$$

From Eqs. (10.57) and (10.59), one obtains

$$\begin{bmatrix} N_x \\ N_u \end{bmatrix} = \begin{bmatrix} A - I & B \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} \Theta \\ I \end{bmatrix} \tag{10.60}$$

The controller K design has the same steps as discussed previously and the system dynamics are dictated by $A - BK$. The full-state feedback is used and the feedback state vector can consist of only measured states or all estimated states through a full-order estimator or part measured states and part estimated states through a reduced order estimator. The N_x and N_u matrices for tracking input are decided by Eq. (10.60).

EXAMPLE 10.7

Consider the example of the speed control of the DC motor discussed in the previous examples. The controller gain K is calculated based on $A - BK$ dynamics using the Ackerman's formula as discussed in the regulator controller design example. Only the scaling matrices N_x and N_u need to be determined. Applying Eq. (10.60) and with the help of a computer program, the N_x and N_u matrices for the DC motor example are

$$N_x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$N_u = [4.1]$$

Referring to the control block diagram shown in Figure 10.41, there is only one controlled output or performance output (i.e., the angular speed, ω). Therefore the reference vector consists of only one variable. The control input vector is also only one variable, that is, the armature voltage. Thus,

$$x_{ref} = N_x \cdot r = N_x \cdot \omega_{ref} = \begin{bmatrix} 2 \cdot \omega_{ref} \\ \omega_{ref} \end{bmatrix}$$

This x_{ref} is compared with the fed back state vector x. The error is fed as input to the controller K. The control input u is given as

$$u_k = N_u \cdot r + e_k K = N_u \cdot \omega_{ref} + (N_x \cdot \omega_{ref} - x_k)K$$

The fed back state vector x_k can be directly fed back or obtained from an estimator.

10.10 Controlling Voltage

The previous sections discussed the methods for designing the controllers based on the general control structure as depicted in Figure 10.1. If the plant is an inverter generating an AC voltage, then sensing the AC voltage and feeding back the voltage is generally done with the help of potential transformers. If the plant is a DC–DC converter, then voltage sensing is done using opto-couplers.

The general control structure for the converters is shown in Figure 10.41 where the plant considered is a typical flyback converter. The output voltage V_o is the controlled output and the duty cycle for the power switch is the control input. The output voltage is sensed by a sensing and isolation circuit and fed back to the controller. The fed back voltage is compared with the reference $V_{oref.}$ The error between the reference and the fed back signals is given as input to the PI or PID controller. The controller output is a voltage signal which has to be converted to time information. The voltage-to-time conversion is done by comparing the controller output with a triangular waveform at the switching frequency. The output of the comparator will give a pulse-width modulated (PWM) signal wherein the controller voltage information is converted to pulse-width modulation or duty cycle variation. The pulse signals are passed through the power switch drive circuit (as described in Chapter 2) to control the power semiconductor switch state.

The output voltage sensing may be done in a simple manner by attenuating the output voltage with a resistive divider. However, in that case the output and the primary-side grounds are common and the circuit looses the galvanic isolation property. To retain the output and the primary-side ground isolation, the opto-isolation circuit as depicted in Figure 10.42 is a popular-sensing circuit used in DC–DC converter circuits. The output voltage is used to drive a current through the diode of the opto-isolator. The current through the opto-isolator diode is

$$i_{opto} = \frac{V_o - V_z}{R_{opto}} \tag{10.61}$$

The current i_{opto} through the opto-coupler diode optically drives the opto-transistor Q. If the output voltage V_o increases, then i_{opto} increases leading to greater drive for the opto-transistor Q and the voltage across the

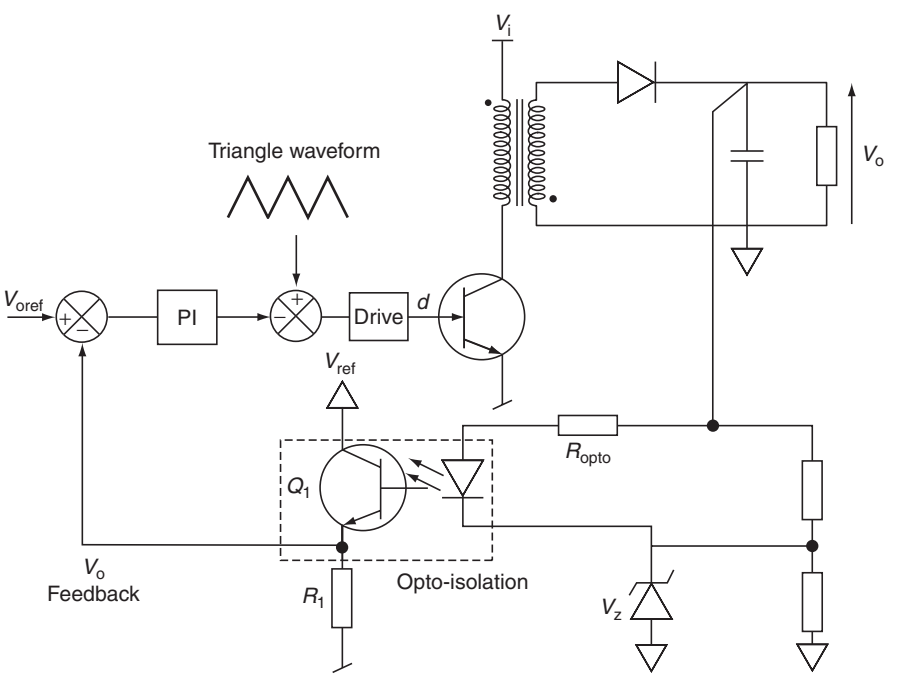


Figure 10.42 Control block diagram for DC–DC converters.

transistor decreases. This implies that the fed back voltage across R_1 increases. Thus, V_o feedback across R_1 is directly proportional to V_o. The voltage reference V_z is used for nominal drive bias and also for temperature compensation of the opto-coupler diode.

The comparator, controller, triangle waveform generation and comparison with the controller output and pulse-width modulation generation are generally available as a single IC. Many such PWM ICs from various manufacturers are available in the market for the designer to make an appropriate selection based on the application. The PWM ICs are mixed signal ICs and the control is performed in the continuous or analog domain. Thus the plant model and the controller design should be done in the *s*-domain. Alternately, the controller and the modulator can be programmed within a microcontroller wherein the controller is in the *z*-domain. In such a case the plant model should be converted into the *z*-domain equivalent by providing the sampling time information and then the controller should be modeled in the *z*-domain. As microcontrollers are more popular and more ubiquitous in usage, the *z*-domain controller design is more relevant in practical applications.

In most applications the DC–DC converters supply multiple loads with isolated grounds and different voltage requirements. In such cases, the transformer consists of multiple secondary windings that supply power to the various loads. From the control viewpoint, as there is only one control input (i.e., the duty cycle of the power switch), only one output can be controlled. Then the question arises as to how the other outputs should be controlled?

For all such applications involving multiple outputs, the voltage corresponding to the highest power output is fed back in the manner described above for controlling it. The other outputs are controlled in one of the following methods:

1. Linear regulator method
2. Non-isolated converter method
3. Coupled inductor method
4. Magnetic amplifier method

Figure 10.43 illustrates the use of a linear regulator to control the non-controlled outputs. The combination of the switched-mode and linear regulator is a very popular method to obtain good quality regulated outputs. The switched-mode converter helps in reducing the input–output differential voltage drop across the linear regulator by providing the necessary voltage scaling. The linear regulator provides the line and load regulation that is required by the output.

Alternately, the linear regulator can be replaced by a non-isolated converter that can be one of the three primary DC–DC converter topologies (Figure 10.44). The power switch of the secondary-side non-isolated converter is controlled by a local feedback loop system that is powered by the output voltage itself.

Another interesting method of controlling the uncontrolled outputs is by coupling the inductors of each of the isolated outputs as indicated in Figure 10.45. Referring to Figure 10.45, let V_{o1} be controlled in the regular manner by feeding back the output voltage through the opto-isolator circuit and let V_{o2} be the uncontrolled output. For the sake of this discussion, consider the secondary diodes as ideal and that the leakage inductances between N_1 and N_2 windings is zero and the coupling between L_1 and L_2 as tight. The following constraint is applied to the above circuit:

$$n_c = \frac{V_{s1}}{V_{s2}} = \frac{N_1}{N_2} = \frac{N_{L1}}{N_{L2}} \tag{10.62}$$

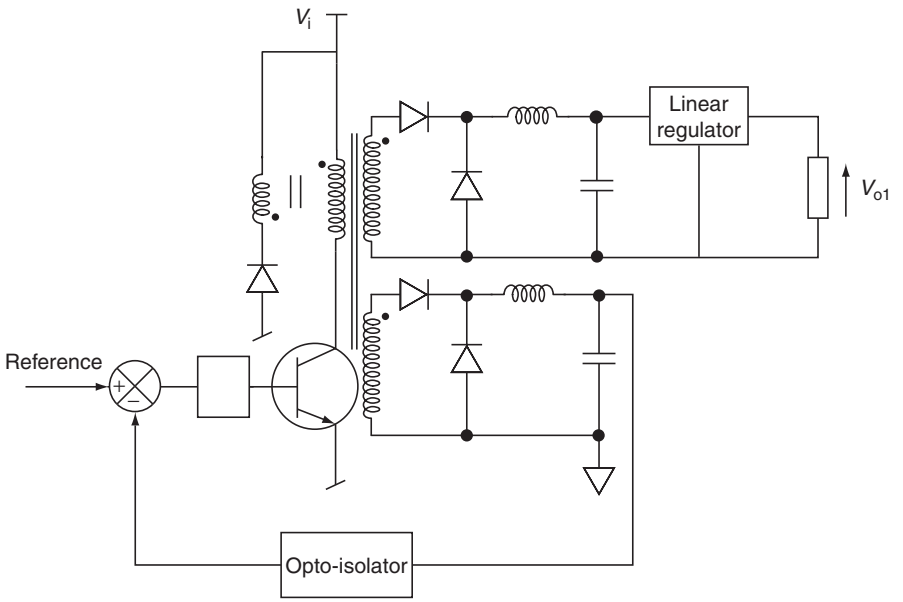


Figure 10.43 | Control of multiple outputs with linear regulators.

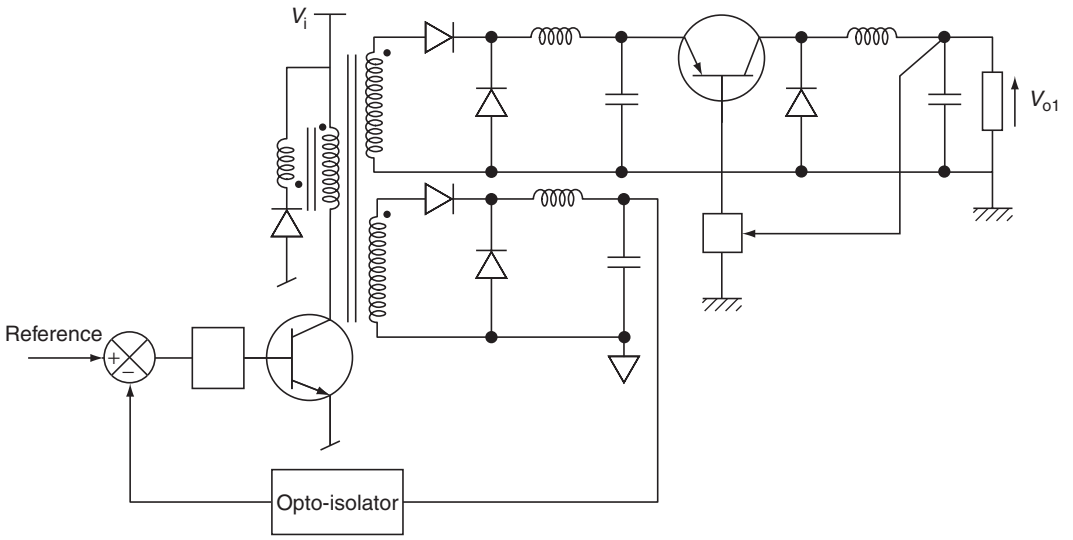


Figure 10.44 | Controlling multiple outputs using non-isolated converters.

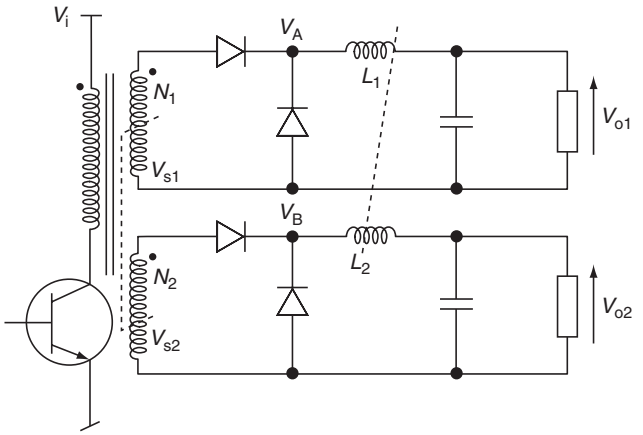


Figure 10.45 | Control of multiple outputs through coupled inductors.

where N_{L1} and N_{L2} are the number of turns of the inductors L_1 and L_2, respectively. Node voltage V_A is equal to V_{s1} during period DT_s and is equal to zero during $(1 - D)T_s$. Likewise, the node voltage V_B is equal to V_{s2} during period DT_s and is equal to zero during $(1 - D)T_s$. From Eq. (10.62), it can be deduced that

$$V_A = n_c \cdot V_B$$

$$V_{L1} = n_c \cdot V_{L2} \tag{10.63}$$

The uncontrolled output voltage V_{o2} is given as

$$V_{o2} = V_B - V_{L2} = n_c V_A - n_c V_{L1} = n_c (V_A - V_{L1}) = n_c V_{o1} \tag{10.64}$$

Equation (10.64) implies that as n_c is a constant, if V_{o1} is controlled then V_{o2} is automatically controlled to n_c times V_{o1}. Any variation in V_{o2} reflects as variation in V_{L2} and hence variation in V_{L1} and V_{o1}. The variation in V_{o1} is controlled by the control loop which in turn brings back V_{o2} to the nominal value. It should, however, be noted that the relationships given in Eq. (10.63) is for an idealized condition where the secondary diodes are ideal and that the coupling between the two secondary windings and the inductors is perfect. In practice this is not possible. The diode drops vary with the load currents and are different in each winding and for different load values. Further, the coupling between the inductors and the transformer secondary windings is not perfect. Therefore the relationship $V_{o2} = n_c V_{o1}$ is not exactly correct. This will cause a certain amount of de-regulation in both the outputs. As a consequence each individual power supply will have to be tuned for coupling to obtain the best regulation. This, therefore, is not a very good method for mass productions.

Another interesting method is the magnetic amplifier method. Actually the name magnetic amplifier is a misnomer as no amplification is involved. It uses a saturable inductor to control the output voltage. The saturable inductor L_s is included in the uncontrolled output circuit as shown in Figure 10.46. A controlled current $i_{\phi R}$ is injected at node B to reset the inductor L_s as indicated.

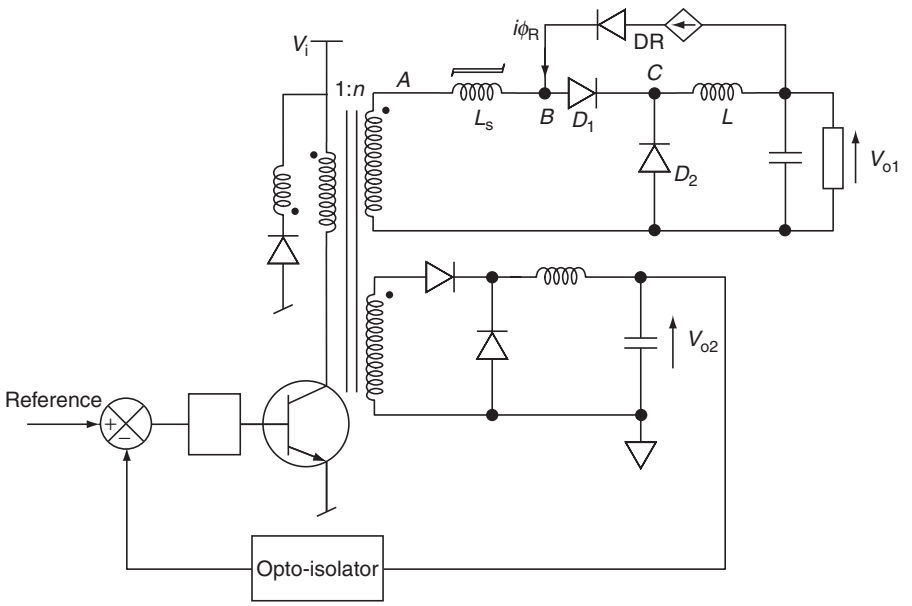


Figure 10.46 | Control of multiple outputs using magnetic amplifier.

Operation

The operation of the magnetic amplifier can be understood by referring to the waveforms shown in Figure 10.47. Consider the steady-state operation wherein the forward-converter switch is OFF and the output inductor L is freewheeling through the diode D_2. The voltage across the secondary is such that the dot end is negative making D_1 reverse-biased and OFF. Now when the forward-converter primary-side switch is turned ON, the dot ends of the transformer windings are positive with respect to the non-dot ends. The inductor L_s offers high impedance and does not allow the voltage to pass immediately to the node C. All the secondary winding voltage drops across L_s. The output inductor L continues to freewheel through D_2. A current flows through the inductor L_s through the path starting from A–L_s–B–D_1–node C–the non-dot end of the secondary winding. The current flows from node C to the non-dot end of the secondary through the diode D_2. As long as this current is less than the freewheeling current of the output inductor L, the diode D_2 is forward-biased and conducts. Due to the current flowing in the inductor, the flux within the core of L_s increases at the rate of nV_i / N_{Ls} where N_{Ls} is the number of turns of the saturable inductor winding. The saturable inductor saturates once the flux in the core reaches ϕ_{sat} as indicated in Figure 10.47(b). Once the flux in the saturable-core saturates, the inductor cannot support any voltage and offers zero impedance. The secondary voltage is now passed on the node C which will turn OFF D_2 and the output inductor L magnetically starts charging up. The node voltage C waveform is shown in Figure 10.47(c).

At the onset of the $(1 - D)T_s$ period, the forward-converter primary-side switch is turned OFF. The non-dot ends are positive with respect to the dot ends. The output inductor starts freewheeling through D_2 and D_1 is OFF. The saturable inductor core is now reset by allowing a current $i_{\phi R}$ to flow through it in the path V_o–B–A–non-dot end–negative of output. The reset current flowing through the inductor L_s sets up an mmf

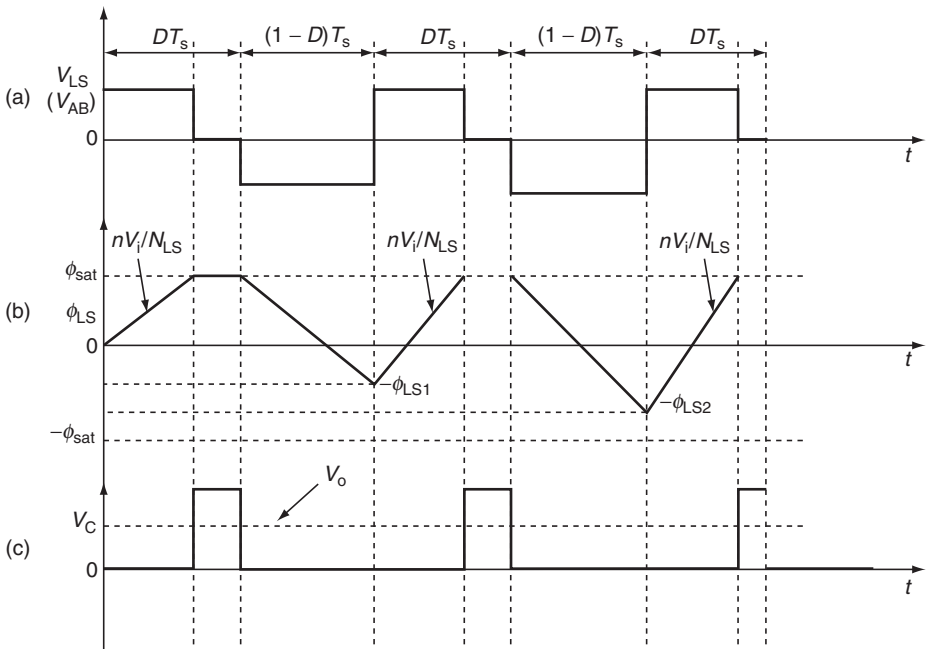


Figure 10.47 | Waveforms explaining the operation of magnetic amplifier.

($N_{Ls} i_{\phi R}$) that resets the core. The value of the current $i_{\phi R}$ decides the extent to which the saturable inductor core gets reset. If $i_{\phi R}$ is small then the core can get reset to a flux value of $-\phi_{Ls1}$ as indicated in Figure 10.47(b). If $i_{\phi R}$ is high then the core can get reset to a flux value of $-\phi_{Ls2}$ as indicated in Figure 10.47(b). The depth to which the L_s core gets reset determines the time taken for the core to saturate in the immediate next DT_s period. As the rate of rise of the flux in the core during the DT_s period is fixed and equal to nV_i / N_{Ls}, the time taken for the core to saturate is wholly dependent on the depth to which the core is reset in the previous cycle. As can be observed from the node C waveform of Figure 10.47(c), if the core reset is not deeply in the negative side as shown for $-\phi_{Ls1}$, the inductor reaches the saturation during the DT_s period in a shorter time as compared to the $-\phi_{Ls2}$ reset value. This implies that if the core is reset to a greater depth in the negative direction, then voltage pulse width available at node C to charge the output inductor is smaller and vice-versa. *Thus larger $i_{\phi R}$ core reset current will try to reduce the output voltage and a smaller $i_{\phi R}$ core reset current will try to increase the output voltage.*

Figure 10.48 gives the practical implementation of the magnetic amplifier. The core reset current $i_{\phi R}$ is obtained by the transistor circuit. The output voltage is compared with a reference and the error is used to drive the PNP transistor Q. The resistor R_e along with Q controls the reset current $i_{\phi R}$. If the output V_o increases, the error at the output of the op-amp reduces. This will increase the drive for the PNP transistor Q which drops less emitter–collector voltage across it. The drop across R_e will increase, which will increase the reset current $i_{\phi R}$. An increased value of $i_{\phi R}$ will reset the core deeper in the negative direction thereby trying to reduce the output voltage. A reduction in the output voltage will in a similar manner lead to reduced $i_{\phi R}$ which will try to increase the output voltage.

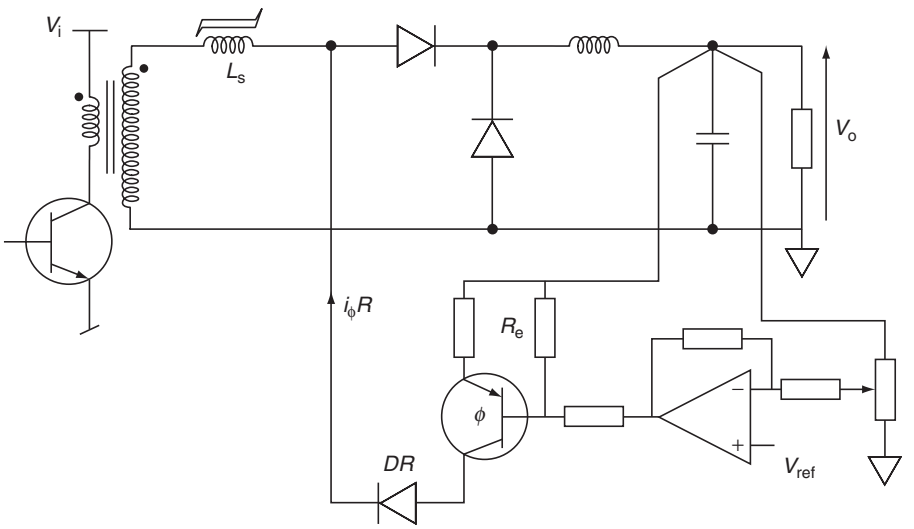


Figure 10.48 | Magnetic amplifier circuit implementation.

10.11 Controlling Current

In most applications, generally the output voltage is the controlled output. However, there are some applications where the input current or the switch current or the output current is controlled. Typical current control applications are

1. current-controlled DC–DC converter or current programmed converter;
2. boost converter based unity power factor converter;
3. single-phase and three-phase front-end converter for reactive power compensation.

These topologies will be discussed to get a flavor for controlling the current instead of the voltage.

Current-Controlled Converter

In this method of control, the DC–DC converter is operated with constant frequency but current-limited pulse width. The turn-ON times of the power switch are clocked periodically and the turn-OFF times are determined by the time at which the current through the switch reaches a reference current limit. This method can be applied to any of the single switch, dual switch or the derived forward converters. This involves inner fast current control loop as shown in Figure 10.49. The advantages of current control are

1. The power semiconductor switch is turned OFF when the switch current reaches a reference current limit. This implies that there is inherent over-current protection for the entire converter.
2. Several similar converters may be connected in parallel to share the load. This is possible as the current through the power switch of each converter is limited by the reference current limit, each converter will deliver only the proportion of power that it is programmed for by its reference current limit.
3. The peak inductor current is determined by the reference current limit that limits the maximum current through the power switch. Thus, the maximum energy-storing capability of the inductor

is determined and as overloads cannot occur in this type of control, the inductor saturation is eliminated.

4. As the power switch currents are limited, the problem of flux walking that is discussed in the case of push–pull, half-bridge and full-bridge topologies does not exist. If there is a mismatch in the ON-state drops of the two sets of switches, then the transformer flux starts drifting towards a saturation limit. As the flux nears the knee of the flux curve with respect to the magnetization or mmf, the currents in the power switches increase which will turn OFF the switch. This will automatically change the duty cycle and will maintain the volt–second balance across the transformer even in the face of mismatched set of power switches. This implies that one need not use the flux-walking capacitor in this mode of control.

Figure 10.49 gives the control schematic of the current programmed control wherein an inner fast current control loop is incorporated in addition to the outer voltage control loop. The DC–DC converter is a typical forward converter. The switch current is sensed using a unidirectional current transformer (CT). The output of the CT is amplified and compared with the output of the voltage controller V_c which sets the reference current limit. At the start of every period, a clock generates a pulse that sets the *S–R* flip–flop which turns the forward-converter power switch ON. The switch has a linearly increasing pattern due to the reflected current from the output inductor. When the switch current reaches the reference current limit V_c, the comparator will generate a negative pulse that will reset the flip–flop and turn OFF the forward-converter power switch.

Slope Compensation

In current programmed control, there is an instability that is introduced due to the inner current control loop for duty cycle operation greater than 0.5. Consider the current flowing in the output inductor at the

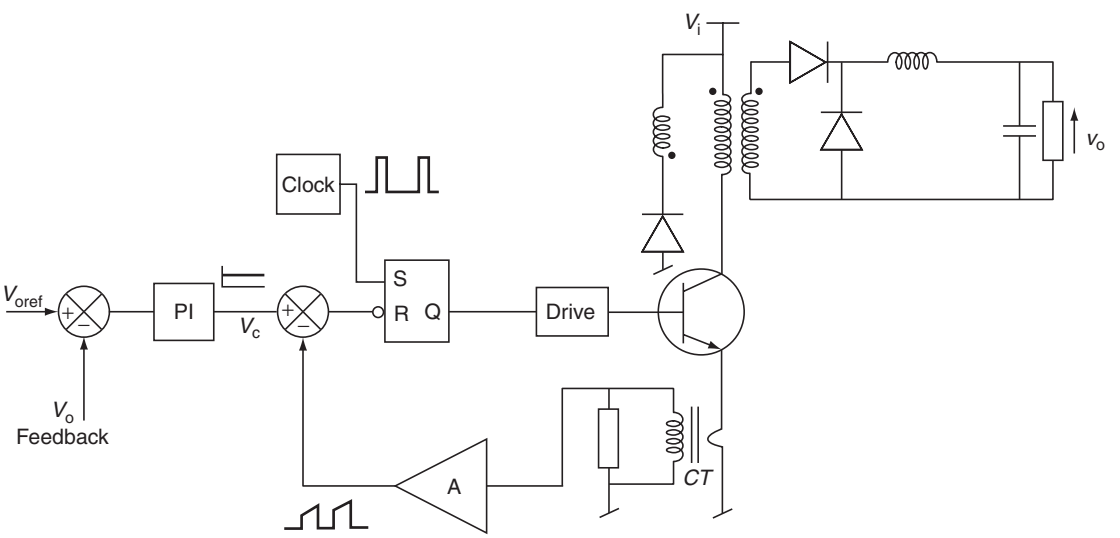


Figure 10.49 | Block schematic of current programmed control.

steady-state condition. The current through the inductor has a linear rise and fall with the slopes dictated by the voltage across the inductor during DT_s and $(1 - D)T_s$ periods. Figure 10.50(a) illustrates the instability problem with the help of the inductor current in steady state.

Consider a forward converter with transformer turns ratio of 1:1 such that the switch current relates directly to the inductor current for the sake of this discussion. However, this discussion is valid for any turns ratio. In Figure 10.50(a), the reference current limit waveform is a constant. During the DT_s period the inductor current and the switch current are same wherein the current rises with a slope of m_1. When the switch current reaches V_c, the switch is turned OFF and the inductor current falls with a slope of $-m_2$. Thus in the steady state,

$$m_1 = \frac{\Delta i_L}{DT_s}; \quad m_2 = -\frac{\Delta i_L}{(1-D)T_s}; \quad \left|\frac{m_2}{m_1}\right| = \frac{D}{1-D} \tag{10.65}$$

Consider a perturbation δi_{L0} in the inductor current at the beginning of the period. The perturbed inductor current follows the dotted lines shown in Figure 10.50(a). At the end of one period, the resulting perturbation is δi_{L1}. This becomes the starting perturbation for the next cycle and so on. At the end of the kth period, the resulting perturbation is δi_{Lk}. From Figure 10.50(a), it can be deduced that

$$\delta i_{L1} = \left|\frac{m_2}{m_1}\right| \delta i_{L0} = \frac{D}{1-D}\delta i_{L0} \tag{10.66}$$

At the end of the kth period, the perturbation δi_{Lk} is given as

$$\delta i_{Lk} = \left(\frac{D}{1-D}\right)^k \delta i_{L0} \tag{10.67}$$

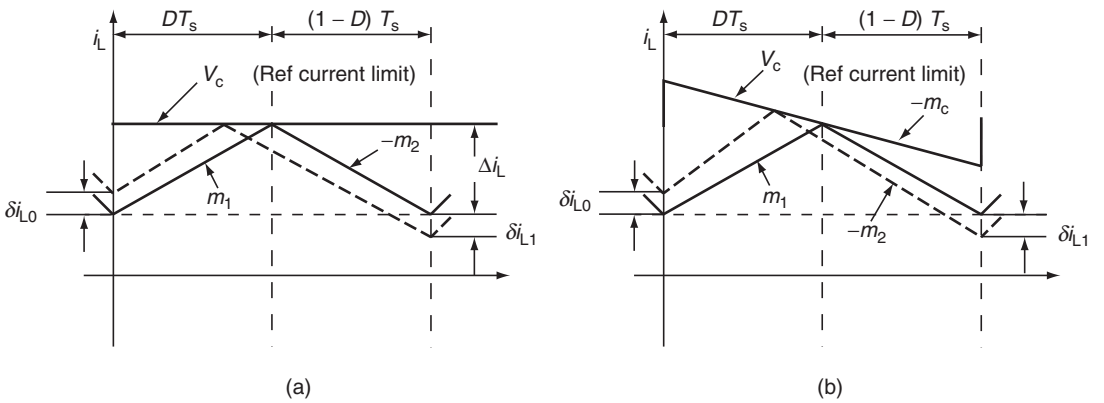


Figure 10.50 (a) Disturbance propagation with constant reference current limit; (b) disturbance propagation with inverted ramp reference current limit.

The term $[D/(1-D)]^k$ grows for $D \geq 0.5$ causing the instability problem. Therefore, for a constant reference current limit, the maximum duty cycle should be limited to 0.5.

The instability problem can be solved by introducing an inverted ramp for the control voltage V_c that determines the reference current limit as indicated in Figure 10.50(b). The current limit voltage V_c is an inverted ramp voltage with a slope of $-m_c$. From Figure 10.50(b), it can be deduced that

$$\delta i_{Lk} = \left(\left| \frac{m_2 - m_c}{m_1 + m_c} \right| \right)^k \delta i_{L0} \tag{10.68}$$

From Eq. (10.68), it can be observed that the initial perturbation δi_{L0} can be made to decay by an appropriate choice of $-m_c$ even for duty cycles beyond $D = 0.5$. In particular, if m_c is chosen equal to m_2, the perturbation decays within one period. Thus, incorporating a stabilizing ramp in the control voltage enables the inner current control loop to be stable. This incorporation of the stabilizing ramp in the control voltage is called *slope compensation*. The modified control block diagram with the slope compensation incorporated is shown in Figure 10.51.

Unity Power Factor Converter

Another application that is very common is the unity power factor converter. In Chapter 3, the rectifier–capacitor filter circuit is discussed. The rectifier–capacitor filter circuit is the most popular method of obtaining a DC voltage from the AC mains. However, the input current is of pulsed nature and as a consequence

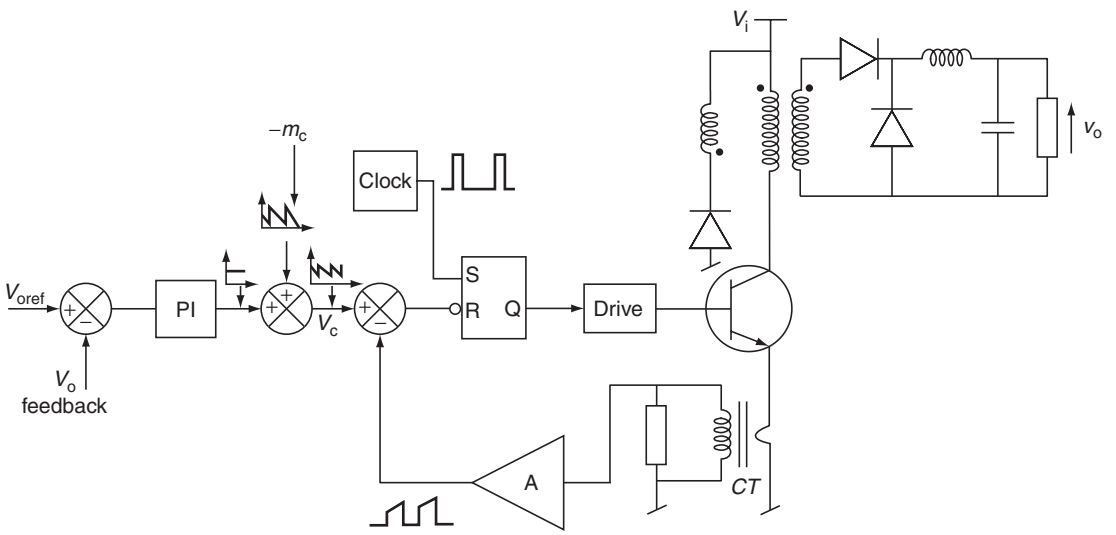


Figure 10.51 | Block schematic of current programmed control with slope compensation.

the power is very poor. To improve the power factor, a boost converter can be used as shown in Figure 10.52(a). The input stage of a boost converter has an inductor that has a filtering effect on the line current. The power semiconductor switch of the boost converter is controlled in a manner such that the input current is in-phase with the rectified source voltage to achieve unity power factor operation. Figure 10.52(b) shows the rectified source voltage that appears at the input to the boost converter. The requirement for the input current of the boost converter is also shown. The input current of the boost converter should be in-phase with the input voltage as shown in Figure 10.52(b), to achieve unity power factor.

The control block diagram of the boost converter for unity power factor operation is shown in Figure 10.53(a). The objective of the controller is to control the input current of the boost converter. The input current i_i of the boost converter is fed back and compared with the reference i_{iref} that has a rectified sinusoidal waveshape such that it is in-phase with the boost input voltage V_i. The error is passed to a PI controller whose output is compared with a triangular signal to generate the PWM signal. The generated PWM is passed through the power switch drive circuit to turn ON and OFF the boost converter switch. The PI controller ensures that the error at its input is zero, which implies that i_i tracks i_{iref}.

Reference Current Generation

The generation of a proper tracking reference i_{iref} is important in this topology. i_{iref} should have a waveshape that follows the waveshape of the rectified source voltage. However, the amplitude of i_{iref} is determined by the output power requirement. Consider the input source voltage

$$v_s = V_m \sin \omega t$$

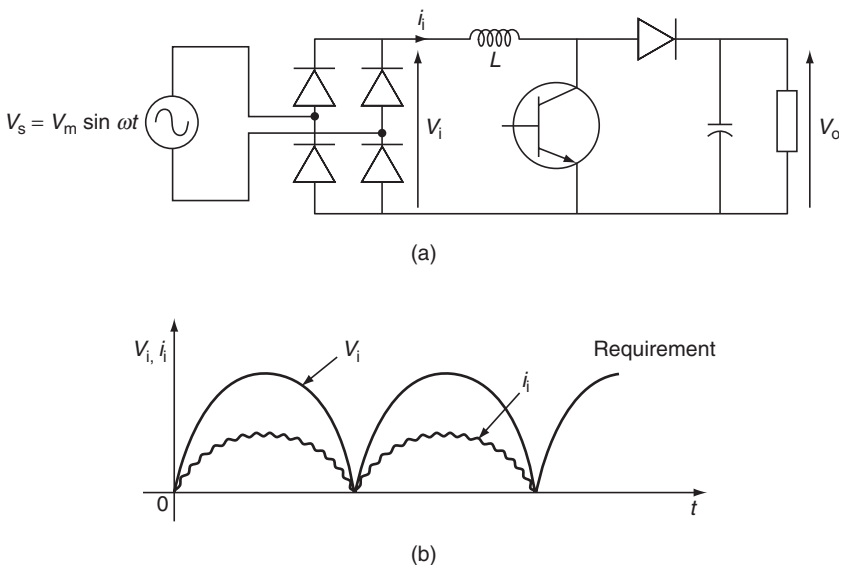


Figure 10.52 (a) Boost converter used as power factor improvement circuit; (b) required input current for unity power factor.

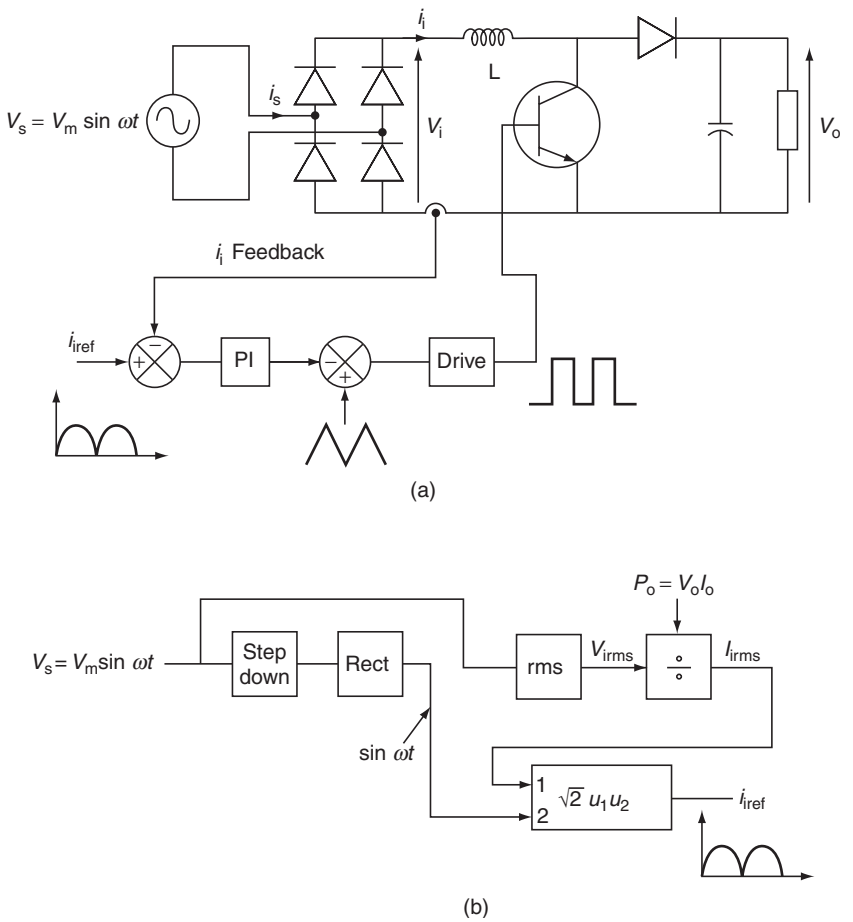


Figure 10.53 (a) Block diagram of the control scheme for power factor improvement; (b) reference current generation.

This v_s is normalized and rectified to obtain

$$v_n = \frac{v_s}{V_m} = \sin \omega t \tag{10.69}$$

Equation (10.69), which is the normalized and rectified source voltage waveform, defines the reference current waveshape. The amplitude is determined by the output power. If P_o is the output power, then

$$P_o = V_o I_o \tag{10.70}$$

Assuming 100% efficiency of the boost converter,

$$V_{irms} I_{irms} = P_o = V_o I_o \tag{10.71}$$

From Eq. (10.71),

$$I_{\text{irms}} = \frac{P_{\text{o}}}{V_{\text{irms}}} = \frac{V_{\text{o}} I_{\text{o}}}{V_{\text{irms}}} \tag{10.72}$$

Equation (10.72) determines the amplitude of the reference current waveshape. From Eqs. (10.70) and (10.72), the reference current is given as

$$i_{\text{iref}} = \sqrt{2}\, I_{\text{irms}} \sin \omega t = \sqrt{2}\, \frac{V_{\text{o}} I_{\text{o}}}{V_{\text{irms}}} \sin \omega t \tag{10.73}$$

The block diagram for the reference current generation is shown in Figure 10.53(b).

Front-End Converter

The unity power factor converter discussed in the previous section is a unidirectional power flow converter. To have a bi-directional power flow wherein the source also sinks power, the front-end converters are used. The topology of a single-phase front-end converter is shown in Figure 10.54(a). The AC source and the two-arm bridge are interfaced by an inductor L as shown. The two-arm bridge is operated as an inverter which generates a voltage V_{inv} such that the current through the inductor or source is in-phase with the source voltage V_{s}. The internal body diodes act as rectifiers to allow the power flow from the source to the capacitive DC bus, V_{dc}.

The control objective here is to maintain the source current i_{s} to be in-phase with the source voltage V_{s}, independent of the nature of the load. The control block diagram is shown in Figure 10.54(b). The voltage to be controlled is the inverter output, V_{inv}. The duty cycle to the four switches are generated that will control the amplitude and the phase of the sinusoidal PWM waveform of the inverter output, such that the source current is in-phase with the source voltage. The voltage V_{inv} or V_{AB} is fed back and compared with a reference $V_{\text{inv}}{}^{*}$ as shown in Figure 10.54(b). The error from the comparator output is passed as input to the controller that tries to make the error zero. The controller output is compared with a triangular waveform to generate the PWM signal. The PWM signal is passed to the drive circuit in order to turn ON and OFF the power switches of the inverter.

Reference Voltage Generation

The inverter output voltage controls the phase and amplitude of the current waveform. Therefore, a proper tracking voltage reference has to be generated that will make the current in-phase with the source voltage independent of the nature of the load. The source current is given as

$$i_{\text{s}} = \frac{1}{L} \int (v_{\text{s}} - v_{\text{inv}})\, \mathrm{d}t \tag{10.74}$$

Let $v_{\text{s}} = V_{\text{m}} \sin \omega t$ be the source voltage. If the source has to only deliver unity power factor current, then the current waveshape should be in-phase with the source voltage. If the peak current is I_{m}, then

$$i_{\text{s}} = I_{\text{m}} \sin \omega t \tag{10.75}$$

Substituting for v_{s} and i_{s} in Eq. (10.74), the required inverter output voltage to achieve unity power factor is

$$v_{\text{inv}}^{*} = \left(v_{\text{s}} - L \frac{\mathrm{d}i_{\text{s}}}{\mathrm{d}t} \right) = V_{\text{m}} \sin \omega t - \omega L \cdot I_{\text{m}} \cos \omega t \tag{10.76}$$

If the inverter output v_{inv} generates a voltage as given in Eq. (10.76), the source current is in-phase with the source voltage. The current amplitude is determined from the output power drawn from the source. If the DC power drawn from the DC bus is P_{o}, then for a 100% efficient inverter.

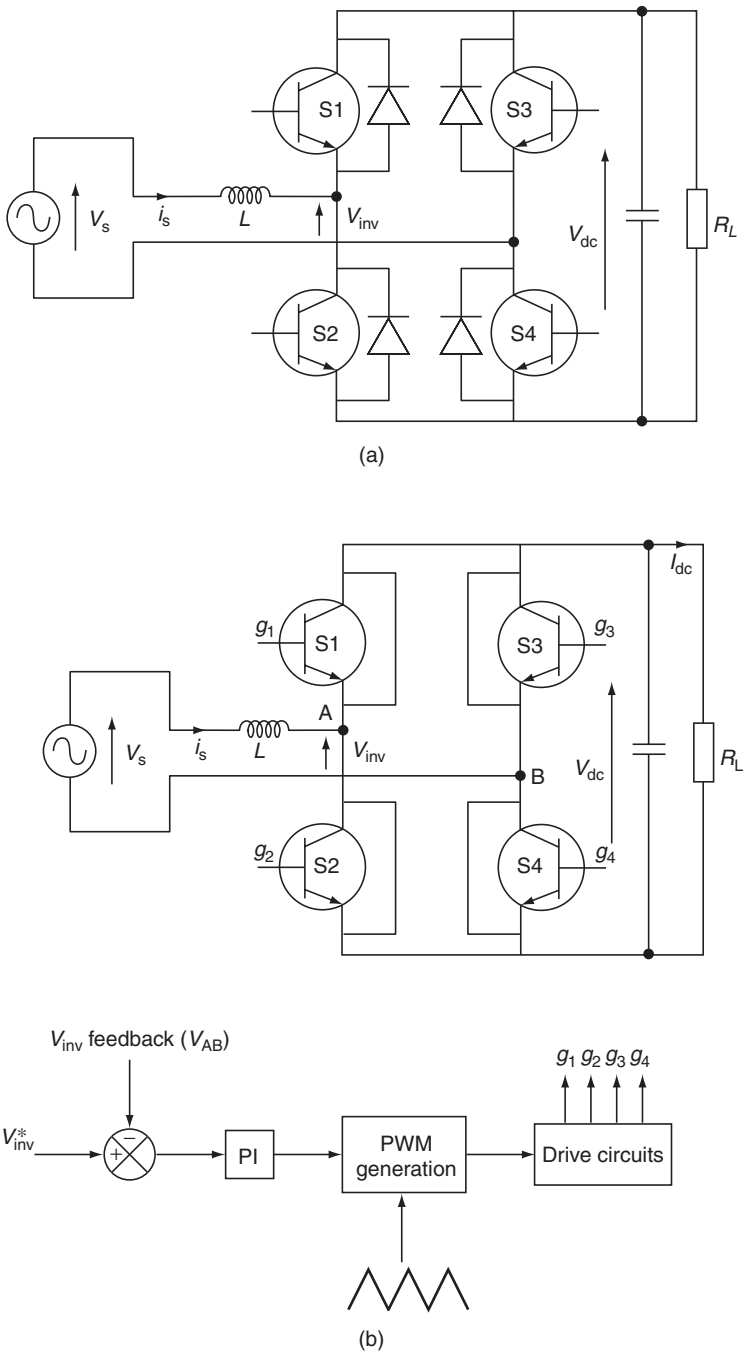


Figure 10.54 (a) Single-phase front-end converter topology for power factor improvement; (b) control schematic for the control of the single-phase converter.

$$P_o = V_{dc} I_{dc} = V_{srms} I_{srms} = \frac{V_m I_m}{2} \tag{10.77}$$

From Eq. (10.77),

$$I_m = \frac{2V_{dc} I_{dc}}{V_m} \tag{10.78}$$

Substituting Eq. (10.78) in Eq. (10.76), one obtains

$$v_{inv}{}^* = V_m \sin \omega t - \omega L \frac{2V_{dc} I_{dc}}{V_m} \cos \omega t \tag{10.79}$$

All the variables and parameters in Eq. (10.79) are known or measurable and hence the reference voltage can be generated to achieve unity power factor operation for the single-phase front-end converter.

Three-Phase Extension

The concepts of the single-phase front-end converter can be extended to the three-phase front-end converter. The three-phase front-end converter topology is shown in Figure 10.55. Here the control objective is to achieve unity power factor operation for the three-phase source. The three-phase currents should be in-phase with the three-phase voltages of the source. Space-vector control method is well-suited for applying to three-phase systems. From the discussion in Section 8.9, the space vector for the three-phase source voltage is given as

$$\underline{v}_s = v_{an} + v_{bn} \cdot e^{i(2\pi/3)} + v_{cn} \cdot e^{i(4\pi/3)} \tag{10.80}$$

Similarly, the space vector for the source phase currents is

$$\underline{i}_s = i_a + i_b \cdot e^{i(2\pi/3)} + i_c \cdot e^{i(4\pi/3)} \tag{10.81}$$

The space vector generated at the output of the inverter at nodes *a*, *b* and *c* is $\underline{v}_{inv}$. The three-phase equivalent circuit is shown in Figure 10.56.

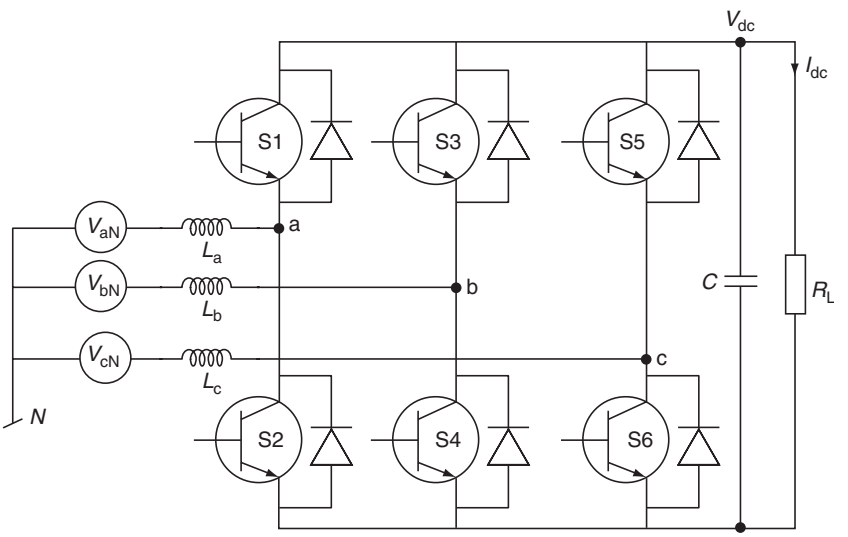


Figure 10.55 | Three-phase front-end converter.

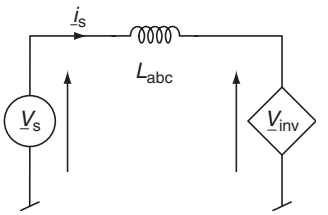


Figure 10.56 | Equivalent circuit of the three-phase front-end converter.

The required reference voltage space vector is given as

$$\underline{v}_{inv}^{\ *} = \left(\underline{v}_s - L\frac{\mathrm{d}\underline{i}_s}{\mathrm{d}t} \right) \tag{10.82}$$

The space vectors for the source voltage and the source current are shown in Figure 10.57 for a general lagging power factor where the current space vector lags the voltage space vector. If the voltage space vector axis is considered as the reference axis, then the reference axis also rotates synchronously with the voltage space vector. With respect to this axis, the current space vector can be resolved into two orthogonal components, i_{sd} and i_{sq}. The i_{sd} component is along the rotating reference axis and the i_{sq} component is perpendicular to the reference axis. For the power factor to be unity, the spatial angle between the voltage and current space vector should be zero. In other words, the current space vector should be along the reference axis. This implies that i_{sq} should be zero. Thus, i_{sq} reference should be set to zero and i_{sd} reference should be set according to the power drawn from the supply, and an inverter voltage space vector $\underline{v}_{inv}$ should be so generated that achieves this objective. All the voltages and currents can be expressed with respect to the reference axis that is aligned along the source voltage space vector and rotating synchronously along with it. This implies that all the current and voltage values with respect to the rotating reference axis will be DC quantities (Figure 10.58).

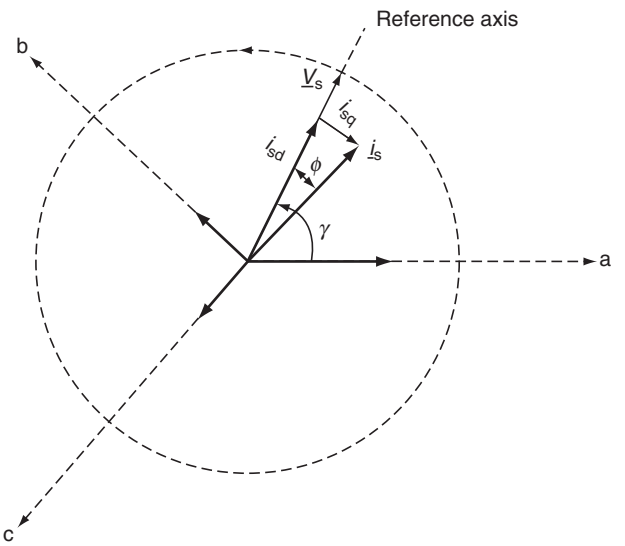


Figure 10.57 | Source voltage and current space vectors.

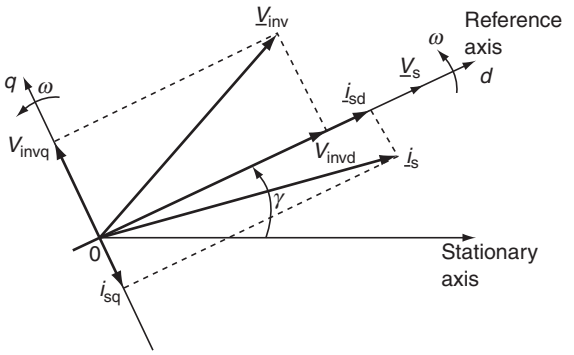


Figure 10.58 Representation of the source voltage, source current and inverter space vectors with respect to the rotating reference axis.

With respect to the rotating reference axis, the inverter voltage space vector $\underline{v}_{inv}$ can be decomposed into the direct and quadrature components as v_{invd} and v_{invq}, respectively. The value v_{invq} should be generated such that i_{sq} is made zero and v_{invd} should be generated such that i_{sd} supplies the power to the load.

All the voltage and current quantities are measured in the stationary axis. Transferring them to the reference axis that is at an angle γ is a two-step process. First the three-phase quantities are converted to two-phase quantities in the same stationary axis. This is called the three-phase to two-phase conversion as discussed in Section 8.9 and given in Eq. (8.89). The second step is to convert the two-phase quantities in the stationary axis to the rotating reference axis by using Eq. (8.92). The variables in the two-phase stationary axis are represented with *a* and *b* subscripts. The variables in the rotating reference frame are represented with *d* and *q* subscripts. If χ is either a voltage or a current variable, then for transferring variables from three-phase stationary axis to the rotating reference axis the following two-step process is used:

1. Three-phase to two-phase transformation in stationary axis:

$$\begin{bmatrix} \chi_\alpha \\ \chi_\beta \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \chi_a \\ \chi_b \\ \chi_c \end{bmatrix} \tag{10.83a}$$

2. Two-phase stationary axis to two-phase rotating reference axis transformation:

$$\begin{bmatrix} \chi_d \\ \chi_q \end{bmatrix} = \begin{bmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} \chi_\alpha \\ \chi_\beta \end{bmatrix} \tag{10.83b}$$

For transferring variables from the two-phase rotating reference axis to the three-phase stationary axis the following two-step process is used:

1. Two-phase rotating reference axis to two-phase stationary axis transformation:

$$\begin{bmatrix} \chi_\alpha \\ \chi_\beta \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \begin{bmatrix} \chi_d \\ \chi_q \end{bmatrix} \tag{10.84a}$$

2. Two-phase to three-phase transformation in the stationary axis:

$$\begin{bmatrix} \chi_a \\ \chi_b \\ \chi_c \end{bmatrix} = \begin{bmatrix} 2/3 & 0 \\ -1/3 & 1/\sqrt{3} \\ -1/3 & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \chi_\alpha \\ \chi_\beta \end{bmatrix} \tag{10.84b}$$

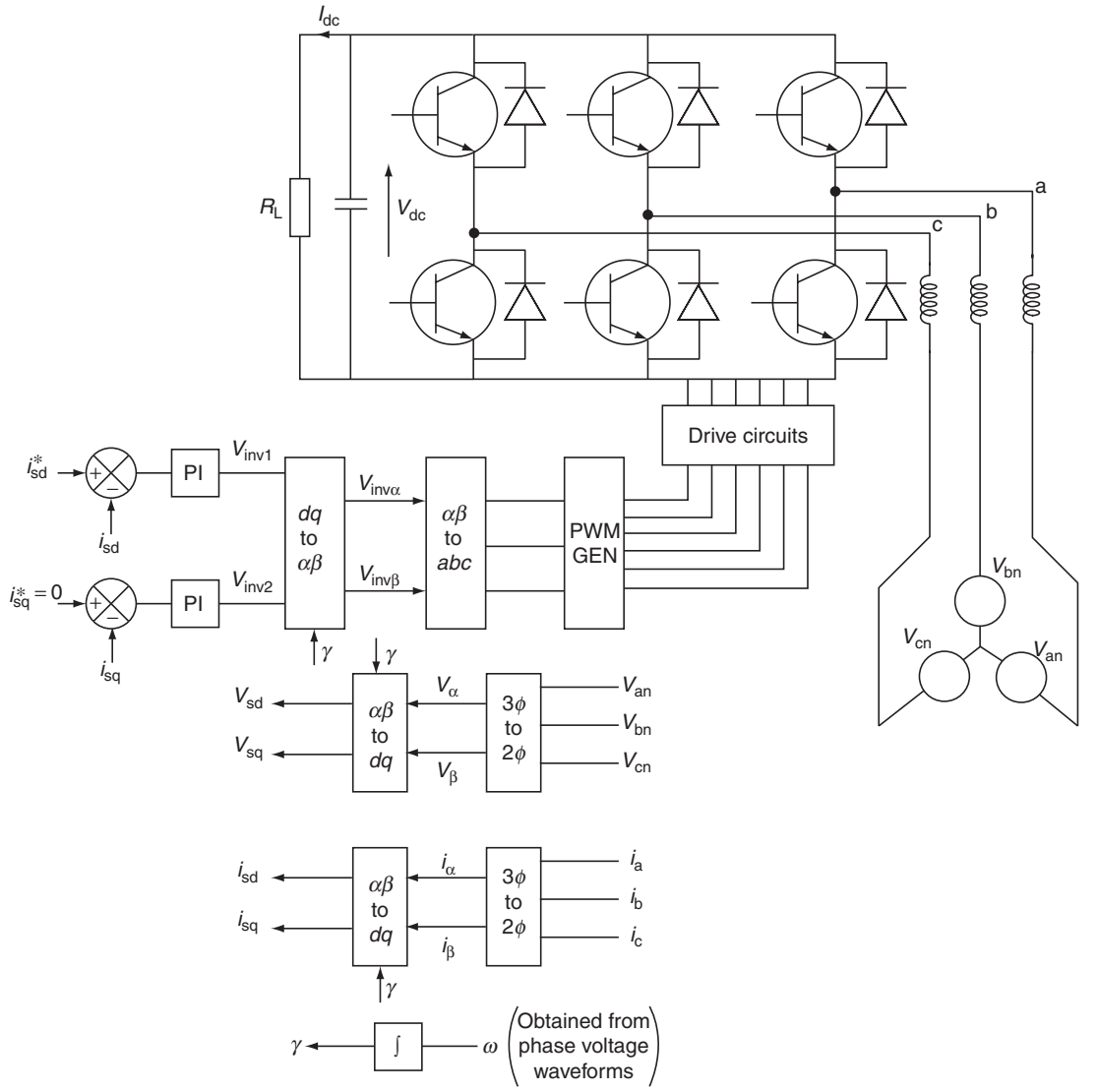


Figure 10.59 | Control block diagram of the three-phase front-end converter.

The angle γ for the axis transformation is obtained from the source voltage waveforms as

$$\gamma = \cos^{-1} \frac{v_\alpha}{\sqrt{v_\alpha{}^2 + v_\beta{}^2}} \tag{10.84c}$$

At the steady state, $\gamma = \omega t$ where ω is the source voltage frequency.

The reference value for the $i_{sq}{}^*$ is zero. The reference value for i_{sd} is obtained from the power delivered to the load. Assuming 100% efficiency of the inverter, the power delivered to the load is

$$P_o = V_{dc} I_{dc} = 3\left(\frac{V_m I_m}{2}\right) \tag{10.85}$$

where V_m is the peak phase voltage and I_m is peak phase current value of the source. In the rotating reference frame, the amplitude of the space vector is 1.5 times the peak amplitude of the phase quantity in the stationary axis due to the transformation. Therefore, in the rotating reference frame,

$$P_o = 3\left(\frac{V_m I_m}{2}\right) = V_{sd} I_{sd} + V_{sq} I_{sq} \tag{10.86}$$

Note that $V_{sq} = 0$ along the rotating reference axis, $V_{sd} = V_s = 3V_m/2$ and $I_{sd} = 3I_m/2$. Therefore, to ensure that the power calculated in the stationary axis is the same as the power calculated in the rotating reference axis, a factor of 2/3 is included to make the power balance. That is

$$P_o = 3\left(\frac{V_m I_m}{2}\right) = V_{sd} I_{sd} + V_{sq} I_{sq} = \frac{2}{3} V_s I_{sd} \tag{10.87}$$

From Eqs. (10.85) and (10.87), the reference value for i_{sd} is

$$i_{sd}{}^* = \frac{V_{dc} I_{dc}}{(2/3) V_s} \tag{10.88}$$

The control block diagram for the three-phase front-end converter control for unity power factor is shown in Figure 10.59.

10.12 Control of Induction Motor

The speed control of induction motor has evolved over the years from the simple stator voltage control wherein the stator voltage is varied to achieve a small speed range. However, the stator voltage control is very inefficient as the flux decreases with decrease in voltage. The generated torque is the product of the flux and the stator current. If the flux decreases with reduction in stator voltage, the generated torque reduces. To produce the same generated torque, the stator current should increase which will also increase the losses. Next in the evolution of the induction motor control is the flux control. Depending on the type of flux control, whether based on the steady-state flux control or dynamic flux control, the scalar (V/f) control and the vector control strategies have evolved.

Scalar Control

The scalar control of induction motor uses the steady-state model of the induction motor. The steady-state model of the induction motor is given in Figure 10.60. The equivalent circuit is similar to that of a transformer. The stator side comprises the stator resistance R_s and the stator leakage $L\sigma_s$. The rotor side comprises the rotor leakage $L\sigma_r$ and the equivalent load is given as R_r/s where s is the slip of the induction motor. The inductance M represents the mutual coupling between the stator and the rotor.

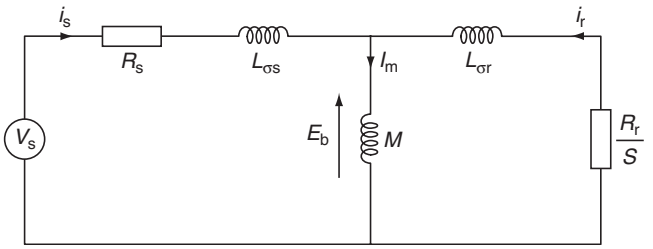


Figure 10.60 | Steady-state model of the induction motor.

The voltage across the mutual inductance M is the back emf which is represented as E_b. If I_m is the magnetizing current responsible for the air-gap flux, then $E_b = \omega M I_m$. The flux linkage in the air gap is given as

$$\phi = MI_m = \frac{E_b}{\omega} = \frac{E_b}{2\pi f} \tag{10.89}$$

From Eq. (10.89), it can be observed that the flux in the core can be maintained constant if the ratio of the back emf (E_b) to the frequency (f) is kept constant. If the flux in the core is maintained as constant, then as the stator voltage is varied the frequency is also varied proportionally. This will give a wide speed range of operation without compromising the torque-generation capability of the induction motor. Figure 10.61(a) gives the relationship between the back emf and the frequency. Upto the base frequency f_b, the back emf and the frequency are linearly related with a constant slope. Beyond base frequency, the stator supply voltage reaches the rated voltage of the induction motor and cannot be increased any further. As only the frequency increases, the flux reduces according to Eq. (10.89). This region beyond base frequency is called the field weakening region.

However, it is not easy to measure the back emf for the induction motor as the equivalent mutual inductance terminal is virtual and not accessible. Therefore, the terminal voltage V_s is considered instead of the back emf. If V_s/f is maintained constant, then the flux within the air gap is reasonably constant. The stator terminal voltage versus the frequency plot is shown in Figure 10.61(b). The terminal voltage V_s differs from the back emf E_b by an amount equivalent to the drop across the stator resistance R_s and the drop across the leakage reactance $\omega L\sigma_s$. The leakage reactance is generally very small compared to the mutual reactance and

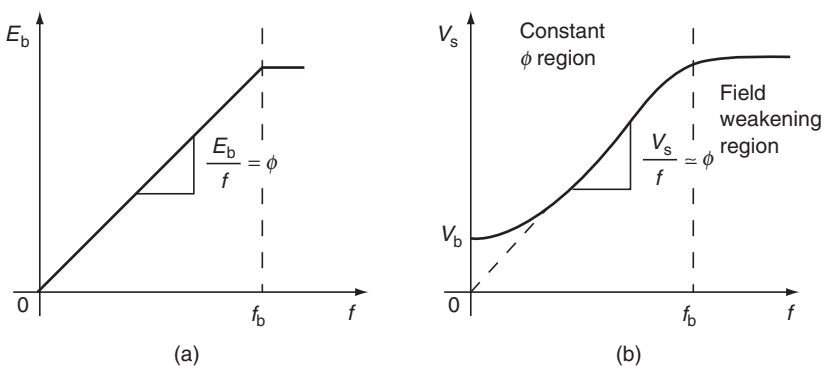


Figure 10.61 | (a) Back emf versus frequency; (b) stator terminal voltage versus frequency.

hence may be neglected. The stator resistance drop is a small fraction of the terminal voltage at high values of V_s, that is, at higher frequency and speed. However, at low frequency, the stator terminal voltage is comparable to the stator resistance drop. As a consequence, the terminal voltage is much different from the actual back emf E_b at low frequencies. To compensate for the stator resistance drop, a small boost to the stator terminal voltage is provided at low frequencies as indicated in Figure 10.61(b). Beyond the base frequency f_b, the terminal voltage is held constant and the field weakens on increasing the frequency.

Open-Loop V/f Control

The control structure of the open-loop *V/f* control of the induction motor is shown in Figure 10.62. The power circuit consists of the mains voltage that is rectified and filtered with a capacitor as shown. The capacitor voltage acts as a DC bus for the inverter. The three-phase inverter supplies the induction motor. The switches of the inverter are controlled by a pulse-width modulator followed by relevant switch drive circuits. The control command is obtained by setting the synchronous frequency ω_s for the induction motor. The required stator terminal voltage V_s is obtained from a look-up table that includes the characteristic as given in Figure 10.61(b).

Integration of ω_s gives the angle γ of the voltage space vector V_s at any given instant of time. The two phase quantities in the stationary axis are obtained as

$$\begin{aligned} V_{s\alpha} &= V_s \cos\gamma \\ V_{s\beta} &= V_s \sin\gamma \end{aligned} \tag{10.90}$$

The two-phase to three-phase conversion is performed as discussed in the previous section and given in Eq. (10.84b). The three-phase stator voltage signal is passed through a PWM generator to drive the switches of the inverter.

Closed-Loop V/f Control

The block diagram for the closed-loop *V/f* control is shown in Figure 10.63. The shaft speed of the induction motor is sensed by a tacho and fed back to the controller. The fed back shaft speed ω_m is compared with reference shaft speed ω_m^*. The error is passed through a PI controller and added with the shaft speed to give the synchronous speed ω_s. The rest of the block diagram is same as that of the open-loop *V/f* control. If the actual induction motor speed increases due to an external disturbance, the speed error goes negative and decreases the slip speed ω_{sl}. This will reduce the synchronous speed ω_s which in turn will reduce V_s. This will have a reducing effect on the actual induction motor shaft speed trying to bring back the induction motor

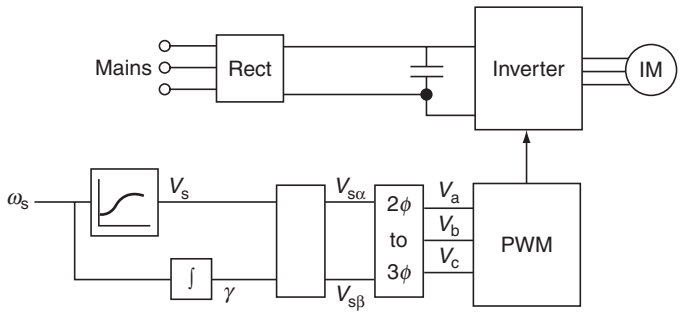


Figure 10.62 | Open-loop *V/f* control.

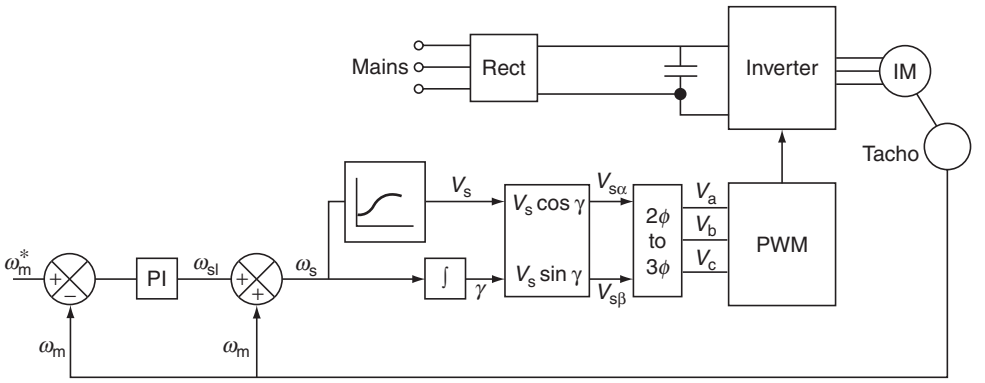


Figure 10.63 | Closed-loop *V*/*f* control.

speed to the set reference speed. A similar control action will take place even when the induction motor shaft speed reduces due to an external disturbance.

Vector Control

In high-performance applications, it is necessary that the torque of the induction motor exhibits a fast response. The dynamics of the electromagnetic torque is shown in Figure10.64(a). Figure 10.64(a) is derived from the generated torque model of Eq. (8.150) that is discussed in Chapter 8. The torque model is based on the rotor flux linkage components and the stator current components in the arbitrary reference frame. The dynamics of the rotor flux linkage components ψ_{rd} and ψ_{rq} shown in Figure 10.64 are evident from the rotor flux state model of the induction motor as given in the state space model of Eqs. (8.128)–(8.130). If the stator current space vector i_s or its components i_{sd} or i_{sq} are changed, then both the flux in the motor and the torque developed by the motor are affected. Thus, in conventional control strategies, there is no way to independently control the flux in the machine and the torque produced by the machine. The torque

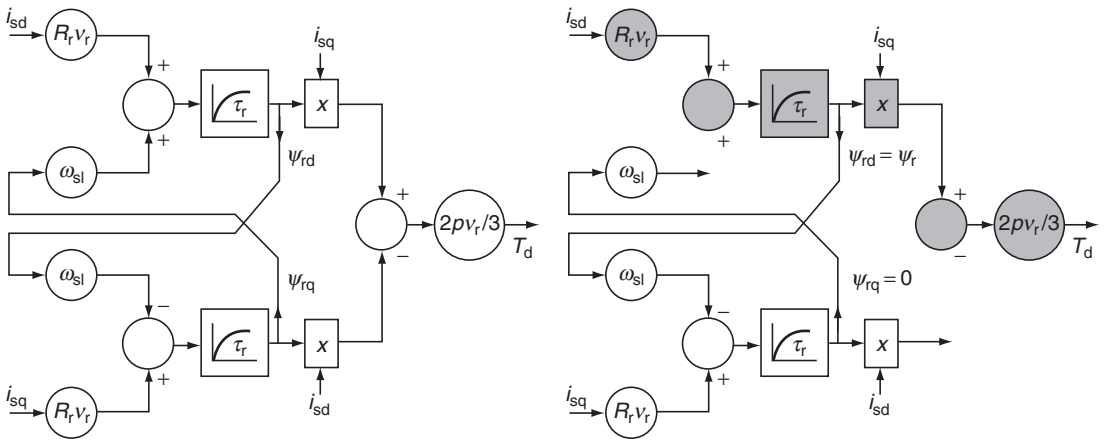


Figure 10.64 | (a) Induction motor torque dynamics; (b) torque dynamics under field orientation.

dynamics will be limited by the dynamics of the flux, which has a large time constant (~300 ms). To improve the torque dynamics, the load torque-producing and flux-producing components of the stator current space vector (i.e., i_{sq} and i_{sd}) must be decoupled. Thus on decoupling, the torque will have dynamics as that of the stator currents (~10 ms) which is generally 20 to 50 times faster than the flux dynamics. The decoupling of the torque-producing and the flux-producing components of the stator current to achieve high dynamic performance in torque is performed by a method called the *vector control or field-oriented control*.

With reference to Figure 10.64, it is evident that any change in the stator current components i_{sd} and/or i_{sq} will affect the rotor flux linkage components ψ_{rd} and ψ_{rq}, with a time constant τ_r. From Figure 10.64, it is clear that if ψ_{rq} is somehow made equal to zero, T_d will depend only on i_{sq} and ψ_{rd} [see shaded portion of Figure 10.64(b)]; and ψ_{rd} will depend only on i_{sd} (without depending on i_{sq}). The drive torque T_d now responds with a time constant τ_r for any deviations in the direct component i_{sd} of the stator current space phasor and will have an almost instantaneous response for any deviations in the quadrature component i_{sq}.

Figure10.65(a) shows the space-vector representation of the stator currents and the rotor flux linkage components with respect to an arbitrary reference frame that is not aligned with the rotor flux linkage space vector. The quadrature component of the rotor flux linkage space vector ψ_{rq} is made zero by aligning the direct axis of the arbitrary reference frame along the rotor flux linkage space vector. When ψ_{rq} is made zero, then from Eq. (8.150) the drive torque T_d is given by

$$T_d = \frac{2}{3} p \cdot \upsilon_r \cdot \psi_{rd} \cdot i_{sq} \tag{10.91}$$

It is evident that if i_{sd} is kept constant, then ψ_{rd} is constant and as a consequence the torque will now have the dynamics of the stator current i_{sq} only. Clearly the i_{sd} component of the stator current can be used to set the magnitude of the rotor flux linkage space vector ψ_r (with $\psi_{rd} = \psi_r$; $\psi_{rq} = 0$) and the i_{sq} component of the stator current can be used to control the magnitude of the drive torque. Thus, by aligning the arbitrary reference axis along the rotor flux linkage space vector, one obtains decoupling between the torque-producing and the flux-producing components of the stator current space vector and as a consequence the drive torque

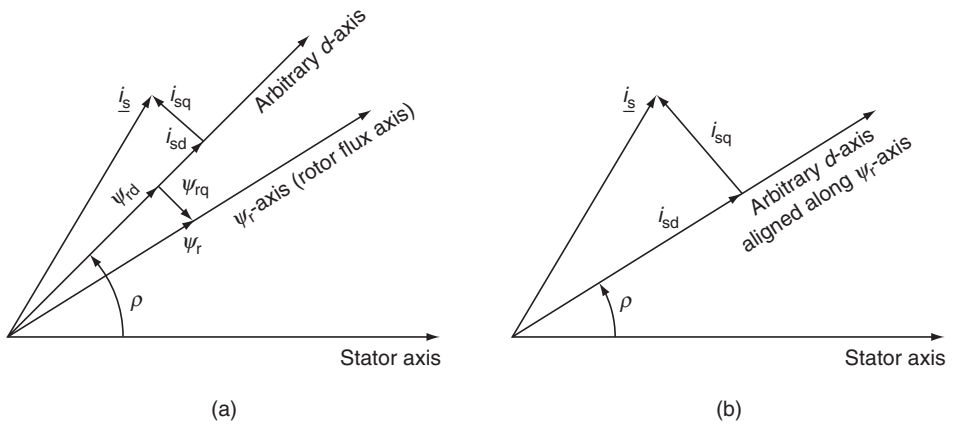


Figure 10.65 (a) Space vectors represented with respect to the arbitrary reference frame, where the direct axis is not aligned along the rotor flux linkage space vector ψ_r; (b) the direct axis of the arbitrary reference frame aligned along rotor flux linkage space vector ψ_r.

dynamic performance is improved. Therefore, this technique of control of the induction motor is referred to as the rotor field-oriented control or simply field-oriented control. In the above scheme, the stator current space vector is controlled by controlling the individual component vectors i_{sd} and i_{sq}. Therefore, this method is also referred to as the *vector control*.

It is interesting to note from Eq. (10.91) that the dynamic performance of a field-oriented controlled induction motor resembles closely to that of the separately excited DC motor which has a similar electromagnetic torque expression. Therefore, it is evident that the field-oriented controlled induction motor or a vector-controlled induction motor can achieve dynamic performances that were in the past achievable only by separately excited DC motors.

It may be noted that one may orient the arbitrary reference axis either along the rotor flux axis (rotor field-oriented control) or the stator flux axis (stator field-oriented control) or the air-gap flux axis (air-gap field-oriented control). All three strategies of field orientation lead to high dynamic performance with respect to torque, but there exists considerable difference in implementing the various field-orientation strategies.

Control of Induction Motor by Field Orientation

The speed control of an induction motor would consist of (a) an inner fast-acting torque control loop to improve the dynamic performance as well as to limit the large transient currents through the power converter and (b) an outer slower acting speed control loop. In the case of the DC motor, the armature current is directly proportional to the electromagnetic torque. For a separately excited DC motor, torque control loop reduces to armature current control with the field excitation being constant. However, in the case of the induction motor, there is no easy way to reduce the torque control to a current control loop unless a component of the stator current that is proportional to the torque is available.

It would be interesting to compare the control structures of the DC motor drive system and that of the field-oriented controlled induction motor drive system. This would enhance the understanding of field-oriented control for induction motors. Figure 10.66 gives the speed-controlled DC motor drive system. Compare this with the speed-controlled induction motor drive system using the vector control strategy as given in Figure 10.67.

In the DC motor drive system as shown in Figure 10.66, the speed reference ω_{mref} is compared with the measured speed ω_m from the tacho. The speed error is passed through a suitable speed controller. The output of the speed controller becomes the torque or the armature current reference i_{aref}. This armature current reference is compared with the measured armature current i_a. The armature current error is passed through a current controller. Likewise the field current reference i_{fref} is compared with the measured field current i_f, and the field current error is passed through a field controller. The field current error is used to modulate the field converter to set the flux level in the DC motor. Similarly, the armature current error is used to modulate the armature converter to control the torque in the DC motor.

The vector-controlled induction motor drive is shown in Figure 10.67. It can be observed that there are a lot of similarities between the DC motor and the vector-controlled induction motor drive systems. The speed controller is similar to the DC motor drive system. The torque control is performed by the quadrature component of the stator current space vector i_{sq}, whereas in the DC motor it is performed by the armature current i_a. The field control is performed by the direct component of the stator current space vector i_{sd}, whereas in the DC motor it is performed by the field current i_f in the separately excited winding. One should note that in the DC motor drive system, the field control portion and the armature control portion are inherently decoupled. On the other hand, in the case of the induction motor there is no such inherent decoupling. The decoupling in the induction motor is achieved by aligning the arbitrary reference frame along the rotor flux linkage

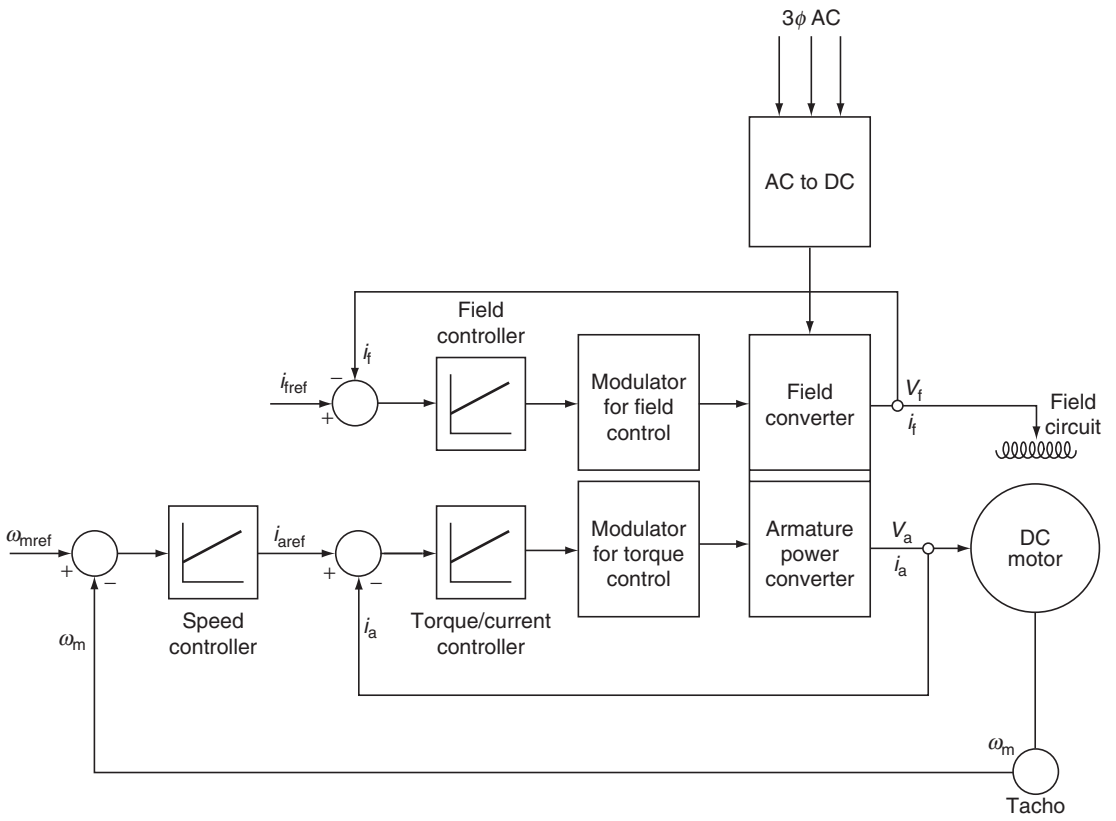


Figure 10.66 | Speed control of separately excited DC motor.

space vector as explained earlier. One should also note that in the control portion of the drive for the induction motor, the variables are in the rotor flux reference frame which is rotating synchronously with the rotor flux linkage space vector. Therefore, all the variables are DC quantities. Thus, a vector-controlled induction motor drive system is similar to a DC motor drive system. Further, the vector-controlled induction motor drive system will give dynamic performances similar to that of DC motor drive system.

For vector control operation of the induction motor, the arbitrary reference frame must be aligned along the rotor flux linkage space vector (or any other flux linkage space vector of interest) at every instant. It is therefore essential that the position of the rotor flux linkage space vector ρ be accurately known at every instant. This knowledge of the rotor flux linkage space vector position can be acquired either by *measuring* the flux directly or by *estimating* the flux from terminal variables (i.e., by indirect means). This leads to two possible control techniques of induction motor:

1. direct field-oriented control;
2. indirect field-oriented control.

Direct Field-Oriented Control The block schematic of a direct field-oriented control (DFOC) of the induction motor is shown in Figure 10.68. The flux measurement can be made using either the Hall sensors

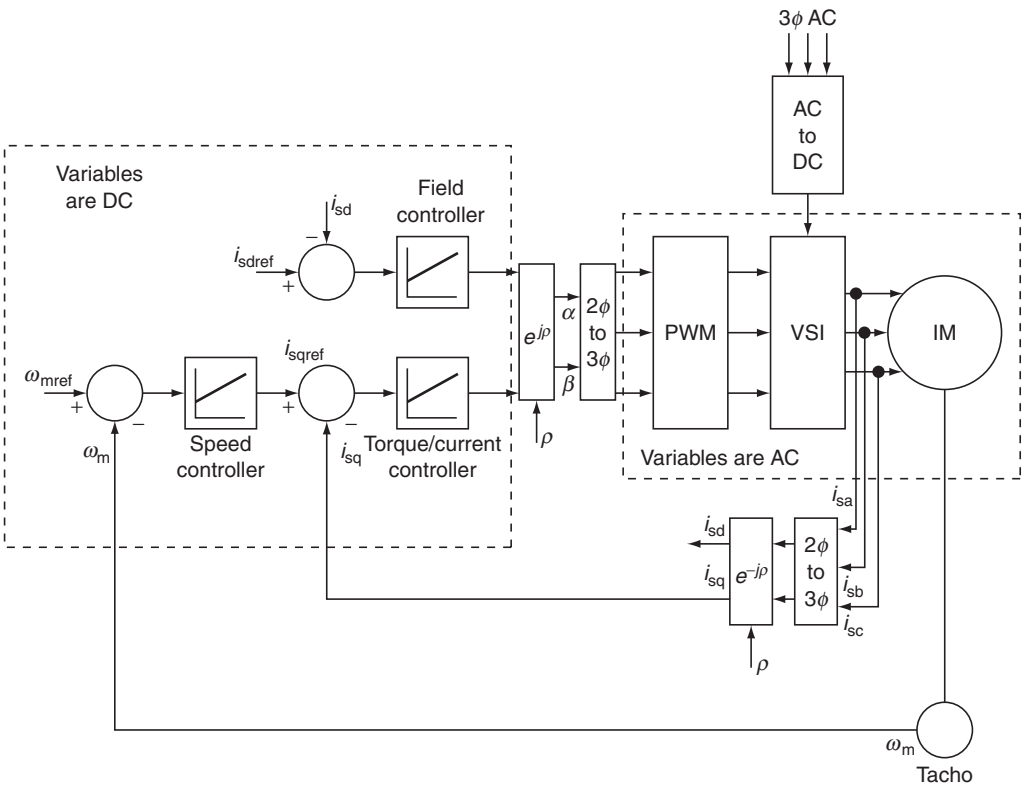


Figure 10.67 | Field-oriented vector control of induction motor.

or the stator search (sense) coils. If the stator sense coils are used, then the voltage sensed from the coils will have to be integrated to obtain the air-gap flux linkages. The measured air-gap flux linkage components are used to calculate the required (rotor, stator or air gap) flux linkage space vector magnitude and position ρ. The value of ρ thus computed is used to align the arbitrary axis along the flux vector to achieve decoupled control of the torque- and flux-producing components of the stator current space vector.

The flux-sensing devices are placed in the air gap of the machine, which will determine the air-gap flux space vector. Any other flux space vector can be calculated as it has an algebraic relationship with the air-gap flux space vector. The air-gap flux sensed by either Hall-effect devices or stator search coils suffers from the disadvantage that a specially constructed induction motor is required. Further, Hall sensors are very sensitive to temperature and mechanical vibrations and the flux signal is distorted by large slot harmonics that cannot be filtered effectively because their frequency varies with motor speed.

In the case of stator search (sense) coils, they are placed in the wedges close to the stator slots to sense the rate of change of air-gap flux. The induced voltage in the search coil is proportional to the rate of change of flux. This induced voltage has to be integrated to obtain the air-gap flux. At speeds below 1 Hz, the induced voltage will be significantly low which would give rise to inaccurate flux sensing due to presence of comparable amplitudes of noise and disturbances in a practical system.

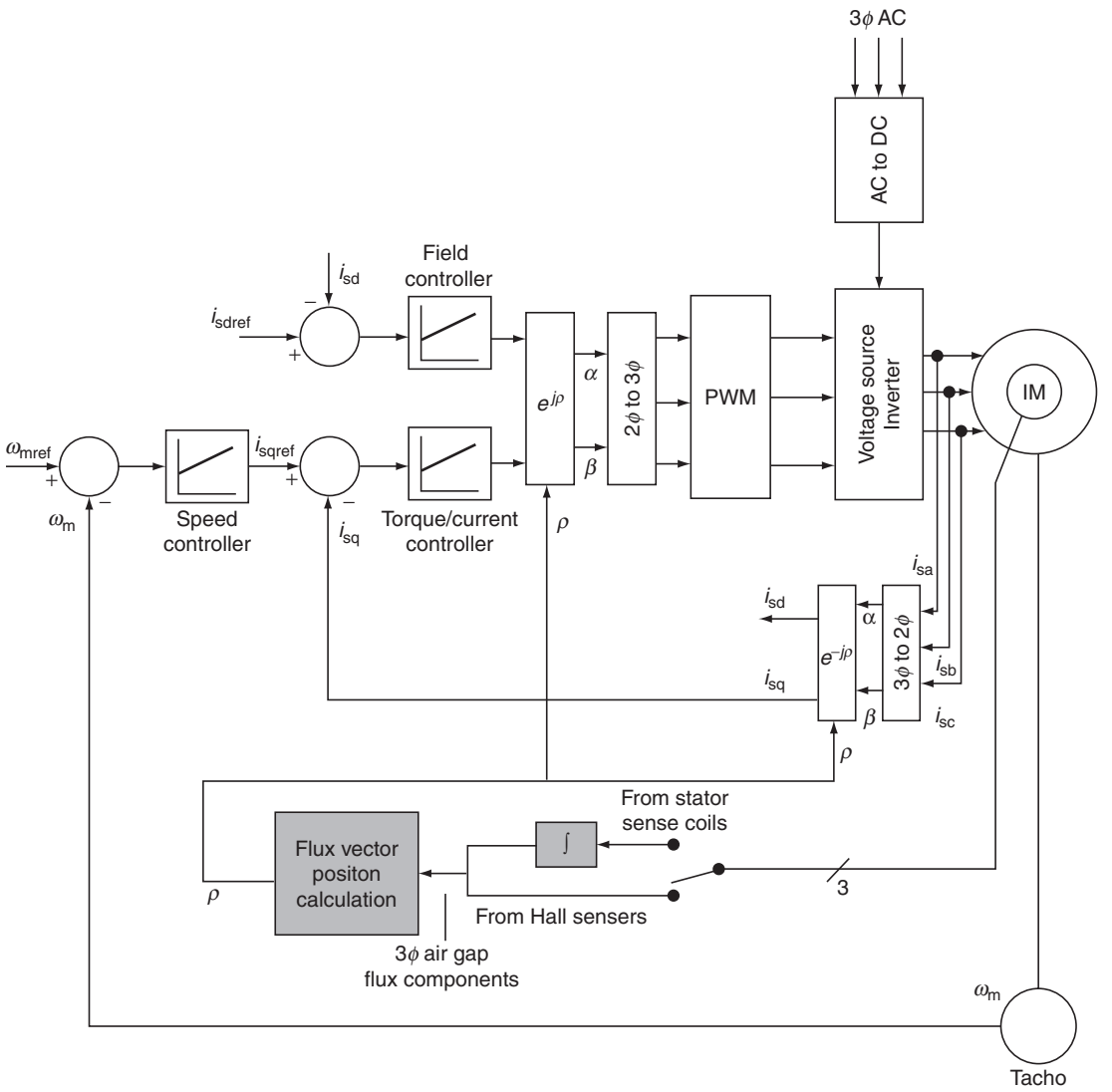


Figure 10.68 | Speed control of direct field-oriented controlled induction motor drive system.

Indirect Field-Oriented Control A flux estimator is used to estimate the required flux linkage space vector magnitude and angular position ρ. The shaft position is usually needed for estimating the flux linkage space vector position. If the shaft transducer is a position encoder, then the position information ε can be directly used. But if the shaft transducer is a speed transducer like a tacho, then the speed has to be integrated to obtain the shaft position. In the case of shaft transducer being a position encoder, the speed feedback is obtained by differentiating the shaft position information.

Indirect sensing of flux space vectors give a more versatile drive system that can be used with standard commercial motors, but this approach would generally result in a more complex control system. Since it is generally desirable to have a scheme which is applicable for all induction motors, the indirect field-oriented control has emerged as the more popular method. In the indirect method of field orientation, the flux linkage space vector is estimated from the motor model. As a consequence, all indirect methods are sensitive to variations in some machine parameter such as the stator or rotor time constants. For example, in the rotor flux-oriented control, the indirect rotor flux estimator is sensitive to the rotor time constant τ_r of the motor. In the case of the stator flux-oriented control, the indirect stator flux estimator is sensitive to the stator time constant of the motor. In the air-gap flux-oriented control, the indirect air-gap flux estimator is sensitive to both the stator and the rotor time constants. Therefore, if the value of the motor parameter varies, the desired decoupling of the flux and the torque components of the stator current space vector is not achieved and this leads to deterioration in the dynamic behavior of the drive system. As a consequence, in indirect field-oriented control, parameter adaptation is essential to overcome the undesirable effects of the parameter variations due to changes in temperature. However, indirect field orientation can be implemented down to zero speed and is therefore suitable for servo drives also.

Flux Estimators for Indirect Field-Oriented Control

The flux estimator is model from the state space representation of the induction motor as discussed in the section on estimators. However, by the knowledge of the plant, specific reduced order estimators can be synthesized. For a rotor field-oriented control, the rotor flux model is deduced from the motor model of Eq. (8.128a).

$$\tau_r \frac{di_{mr}}{dt} + i_{mr} = i_{sd}$$

$$\frac{d\rho}{dt} = \omega_s = \frac{i_{sq}}{\tau_r i_{mr}} + \omega_m \tag{10.92}$$

where $i_{mr} = \psi_r / M$ is defined as the equivalent rotor magnetizing current.

In the practical implementation of the field-oriented control, the induction motor is fed from a voltage source inverter with fast current control loops. This approach is used in high-performance induction motor servo drives for machine tool applications and industrial robots.

In field-oriented control system, the induction motor behaves like a DC machine under both steady-state and transient conditions. Consequently, similar drive control strategies can be employed. Below base speed, the magnetizing current of the induction motor representing the rotor flux magnitude is maintained constant at its maximum possible value to achieve constant torque operation. Above base speed, the flux is reduced, thereby giving the field weakening region or the constant horsepower region of operation.

A typical rotor field-oriented drive system is shown in Figure 10.69. The system consists of an induction motor which is driven by a voltage source inverter. The DC-link to the inverter is obtained from the output of an AC–DC converter which is fed from the three-phase mains. The inverter switching is controlled by the field-oriented controller.

AC–DC converter is the first stage of the power-converter system of the induction motor drive. It consists of a three-phase bridge rectifier followed by a capacitor filter. The output of the capacitor filter acts as the DC-link voltage for the following inverter stage of the power-converter system. This type of AC–DC converter would allow power flow only in one direction, that is, from the mains to the load. In the case of the induction motor drive system, when the induction motor is in the generating mode, the power will flow from the induction motor to the inverter DC-link source. Thus, suitable bi-directional front-end converters which put back the power from the DC-link to the mains have to be used.

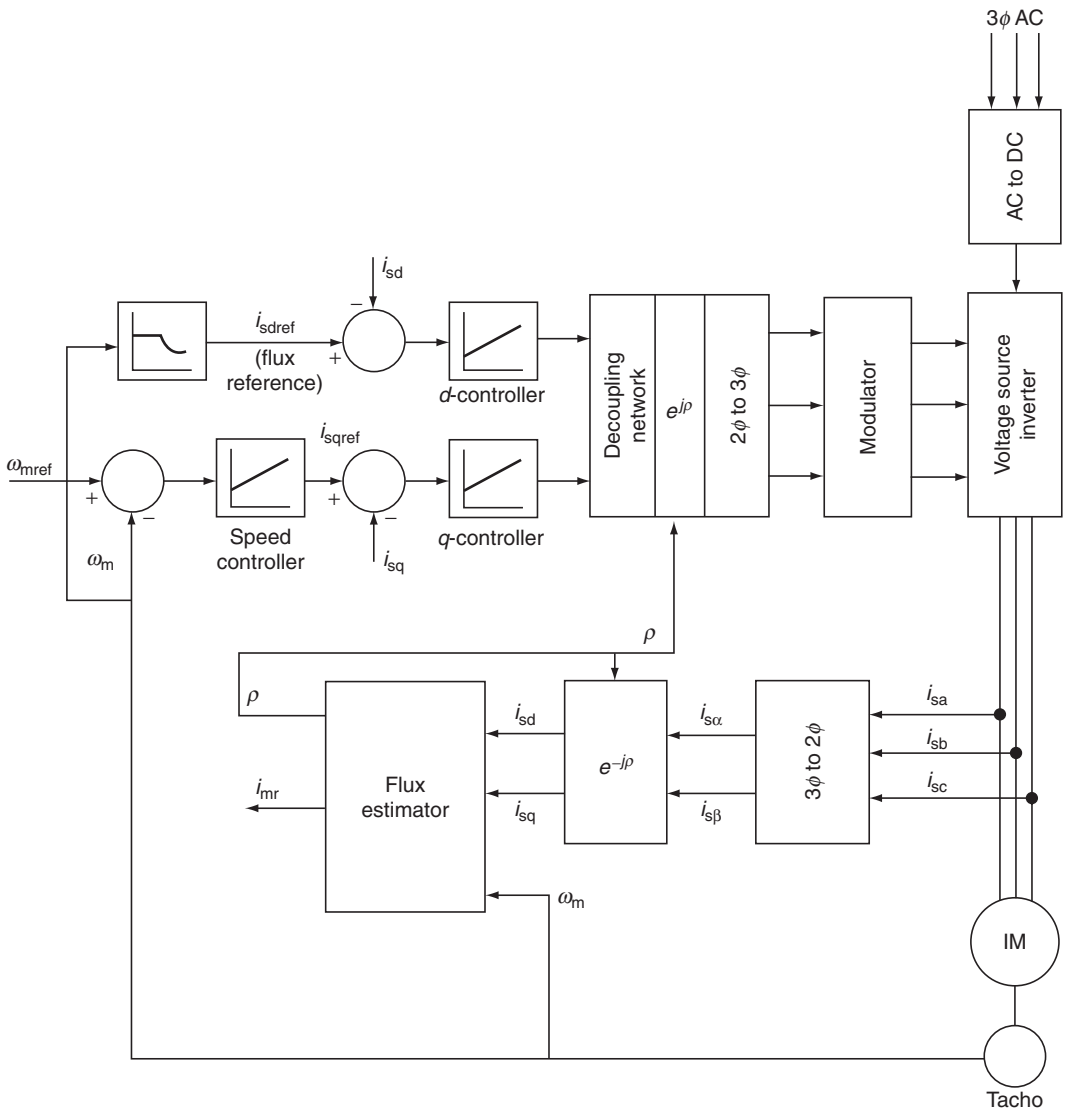


Figure 10.69 | Block schematic of indirect field-oriented control of the induction motor.

The inverter normally consists of six power switches (thyristors, BJTs, MOSFETs or IGBTs) which are connected in the classical bridge configuration. The control signals for the inverter switches are obtained from the modulator block. The six power switch inverter configuration leads to a two-level inverter. Three-level inverter configurations are used for higher powers wherein 12 switches have to be used, four in each inverter leg. Its outputs are three-phase PWM voltages which are fed to the induction motor.

Three-Phase to Two-Phase Transformation (a, b, c to α, β) The three-phase stator currents i_{sa}, i_{sb} and i_{sc} that are measured, are first transformed to an equivalent two-phase system ($i_s\alpha$ and $i_s\beta$) because the induction motor is represented as an equivalent two-phase machine. The three-phase to two-phase transformation ($3\phi - 2\phi$) is carried out in the stator reference frame. The transformation equation is same as Eq. (10.83a). This transformation is a general transformation that can be applied to any variable of the induction motor like the stator voltages, stator currents, flux linkages, etc.

Stator to Synchronous Reference Frame Transformation (α, β to d, q) The two-phase stator currents that are in the stator reference frame are transformed to a synchronous reference frame. The choice of the synchronous reference frame is dependent on the flux along which the orientation is to be performed. If the arbitrary axis is to be oriented along the rotor flux linkage space vector, then the synchronous reference frame would be the rotor flux reference frame and if the arbitrary axis is to be oriented along the stator flux linkage space vector, then the synchronous reference frame would be the stator flux reference frame. If the angle ρ represents the instantaneous position of the synchronous reference frame along which the arbitrary axis is aligned, then the transformation from the stator to the synchronous reference frame is given by Eq. (10.83b). The inputs to this block are $i_{s\alpha}$, $i_{s\beta}$ and the rotor flux position ρ. The outputs of this block are i_{sd} and i_{sq}.

Flux Estimator As already discussed, the magnitude and the position of the flux linkage space vector has to be estimated for indirect field-oriented control. For a rotor flux estimator model, this block has i_{sd}, i_{sq} and ω_m as inputs and ρ and i_{mr} as outputs. The position information ρ is used to perform the $e^{-j\rho}$ and $e^{j\rho}$ transformations [Eqs. (10.83b) and (10.84b)] as shown in Figure 10.69.

Speed and Current Controllers The speed reference ω_{mref} is compared with the fed-back speed ω_m, which is obtained from either a speed transducer or a position encoder or a servo resolver. The speed error thus obtained is passed through a zero steady-state error controller like a PI controller to obtain the command value for the quadrature component of the stator current i_{sqref} (i.e., torque reference) in the synchronous reference frame.

The reference for the direct component i_{sdref} of the stator current space vector in the case of the rotor field-oriented control can be a constant value upto base speed operation of the motor. For the operation of the motor above the base speed, the i_{sdref} decreases in such a manner so as to maintain the power constant, that is, by weakening the field. The command values i_{sdref} and i_{sqref} are compared with the fed back values of the stator currents i_{sd} and i_{sq} in the synchronous reference frame. The current errors thus obtained are passed through PI controllers which form the current controllers of the drive system.

Decoupling Network Under rotor field orientation, the stator current components i_{sd} and i_{sq} are decoupled, and hence the outputs of the current controllers can be used as command values for the current source inverter. However, in the case of a voltage source inverter, the stator voltage command values $-V_{sd}$ and V_{sq} are not decoupled as is clear from the *d–q* model shown in Figure 8.36. Hence decoupling networks are necessary to generate V_{sdref} and V_{sqref} in the synchronous reference frame if a voltage source inverter is used. In the present work, a voltage source inverter is used to drive the induction motor. Therefore, suitable decoupling terms will have to be incorporated to the outputs of the current controllers. Referring to the stator circuit of the *d–q* axes equivalent circuit of the induction motor shown in Figure 8.36, it is observed that the *d*-axis stator circuit loop has a coupling term ($\omega_a\psi_{sq}$) from the quadrature axis and the *q*-axis stator circuit loop has a coupling term ($-\omega_a\psi_{sd}$) from the direct axis. If these coupling terms are not compensated, then the torque and the flux components of the stator current will not be decoupled. Thus feed-forward terms, V_{sd0} for *d*-axis voltage compensation and V_{sq0} for *q*-axis voltage compensation, must be added to the output of the current controllers. V_{sd0} and V_{sq0} are given by

$$V_{sd0} = -\omega_s L_\sigma i_{sq}$$
$$V_{sq0} = \omega_s L_\sigma i_{sd} + \omega_s (L_{ss} - L_\sigma) i_{mr} \tag{10.93}$$

Equations (10.93) gives the decoupling terms that must be added to the outputs of the corresponding current controller outputs to ensure proper decoupling between the direct and the quadrature axis voltages.

PWM Modulator The voltage source inverter operates from a constant DC-link voltage. The variable frequency and variable voltage three-phase output is obtained by suitable pulse-width modulation. A large variety of methods for pulse-width modulation have been developed over the past few decades. Their implementation in the design of an AC drive system depends on the machine type, power level and the semiconductor devices used in the inverter.

Pulse-width modulation strategies for induction motor drives also influence the structure of the current control loops. The current control is performed in the stator reference frame. The references for the d and q axes components (i_{sdref} and i_{sqref}) of the stator current space vector are generated in the synchronous reference frame. They are then transformed to the three-phase stator reference frame as i_{saref}, i_{sbref} and i_{scref}. These three-phase stator current references are compared with the measured three-phase stator current components i_{sa}, i_{sb} and i_{sc} and the corresponding three-phase current errors are passed through a suitable current controller and pulse-width modulator to obtain the switching pattern for the inverter switches. The current controller and the PWM generator can be current hysteresis type or carrier based. The implementation algorithm of three-phase PWM is discussed in Chapter 12.

10.13 Output Feedback

Till now the applications were considered wherein the controller can be the PI or PID type controller or the full-state feedback controller. There is lot of intuition used in the design with PI type controller. However, in the case of full-state feedback controller, once the desired closed-loop pole locations are decided, the controller design is algorithmic. The output feedback method considers only measurable state feedback instead of full-state feedback. All the state components are seldom available for feedback purposes in a practical design problem. As a consequence, the non-available state components are estimated from the available states and the system model, using a reduced order observer. But unfortunately, the state estimates will depend on the accuracy with which the system parameters are known. In the output feedback approach, only the measurable or the available state components of the system are used for feedback control. Further, unlike the full-state feedback, the output feedback control law can be used to design compensators with any desired dynamical structure, thereby regaining much of the intuition of classical controls design.

In many control design applications, it is necessary to have some experience and knowledge which can give indications about the type of the compensator dynamics that would be suitable for the specific application. For example, it may be necessary to augment some feed-forward paths with integrators to obtain a steady-state error as exactly zero. Again, low-pass filters may be required to filter measurements that are noisy. In output feedback control, a dynamic compensator of a prescribed structure may be incorporated into the system.

Consider Figure 10.70 which gives the control structure of a tracker with output feedback wherein the plant is described by

$$x_{k+1} = Ax_k + Bu_k \tag{10.94a}$$

$$y_k = Cx_k \tag{10.94b}$$

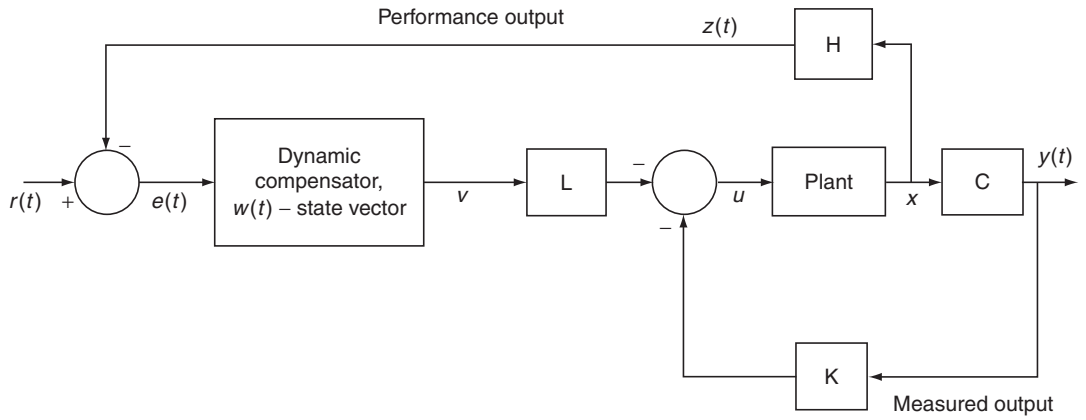


Figure 10.70 | Output feedback control structure.

with state x, control input u and measured output y available for feedback. In addition

$$z_k = Hx_k$$

is a performance output or the controlled output which must track the given reference input r. It is not generally the same as y in all cases.

The dynamic compensator has the form

$$w_{k+1} = Fw_k + Ge_k \tag{10.95a}$$

$$v_k = Jw_k + Me_k \tag{10.95b}$$

with state w, output v and the input equal to the tracking error given by

$$e_k = r_k - z_k \tag{10.95c}$$

Also, F, G, J and M are known matrices chosen to include the desired structure in the compensator. The allowed form for the plant control input is

$$u_k = -Ky_k - Lv_k \tag{10.96}$$

where the constant gain matrices K and L are to be chosen in the controls design step to result in satisfactory tracking of r. This formulation allows for both feedback and the dynamic compensator dynamics. Equation (10.96) is the control law for the output feedback structure shown in Figure 10.70. The augmented system including the plant and the dynamic compensator can be represented using Eqs. (10.94a)–(10.96) as

$$\begin{bmatrix} x \\ w \end{bmatrix}_{k+1} = \begin{bmatrix} A & 0 \\ -GH & F \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}_k + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k + \begin{bmatrix} 0 \\ G \end{bmatrix} r_k \tag{10.97}$$

$$\begin{bmatrix} y \\ v \end{bmatrix}_k = \begin{bmatrix} C & 0 \\ -MH & J \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}_k + \begin{bmatrix} 0 \\ M \end{bmatrix} r_k \tag{10.98}$$

$$z = \begin{bmatrix} H & 0 \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}_k \tag{10.99}$$

and the control input may be expressed as

$$u_k = \begin{bmatrix} -K & -L \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}_k \tag{10.100}$$

The augmented plant description of Eqs. (10.97)–(10.99) can be rewritten by redefining the state, the measured output and the matrix variables such that

$$x_{k+1} = Ax_k + Bu_k + Er_k \tag{10.101a}$$

$$y_k = Cx_k + Fr_k \tag{10.101b}$$

$$z_k = Hx_k \tag{10.101c}$$

with z_k the performance output which is required to track the reference input r. In terms of the new variables, the control law for output feedback is of the form

$$u_k = -Ky_k = -KCx_k - KFr_k \tag{10.102}$$

with the controller gain K to be determined. Using the control law of Eq. (10.68) in Eq. (10.67a), the closed-loop system is

$$x_{k+1} = (A - BKC)x_k + (E - BKF)r_k = A_c x_k + B_c r_k \tag{10.103}$$

Note that in the above formulation of the tracker problem with output feedback, the structure of the controller can be fixed in terms of classical intuition of the controls design. Since the performance specifications of most control systems are given in terms of time-domain criteria and these criteria are closely related to the step response, the reference input r will be considered as a step command with magnitude R. The resulting tracker will work for any arbitrary reference command r.

The dynamics of the augmented system is dictated by $A - BKC$. The design now amounts to the selection of gain K for suitable performance. Since K appears in the feedback and the feed-forward portions of the control input given in Eq. (10.102), this approach amounts to determining K so that both the closed-loop poles and zeros are placed within the unit circle in the z-plane. The Ackerman's formula can be used for determining KC, that is, $KC = \text{acker}\ (A, B, P_c)$, where P_c is the closed-loop pole locations. As the C matrix is known, K can now be obtained by LU decomposition of KC.

In the steady state, $x_{k+1} = x_k = x_\infty$, where x_∞ is the state under steady-state operating conditions given by

$$x_\infty = (I - A_c)^{-1} B_c r_\infty$$

where r_∞ is the reference input under steady-state operating conditions.

The output feedback approach can be understood by applying the structure to an induction motor plant which is discussed in the following section.

10.14 Induction Motor Control with Output Feedback

The output feedback approach will be used for control of induction motor drive system. The induction motor can be split into two dynamical systems as shown in Figure 10.71. One of the systems consists of the speed dynamics. A speed controller is used in the outer loop of the induction motor drive system to account for the speed dynamics. The other system consists of the torque dynamics which is governed by the rotor flux state model of the induction motor. The inner current control loops (i.e., the d-controller and the

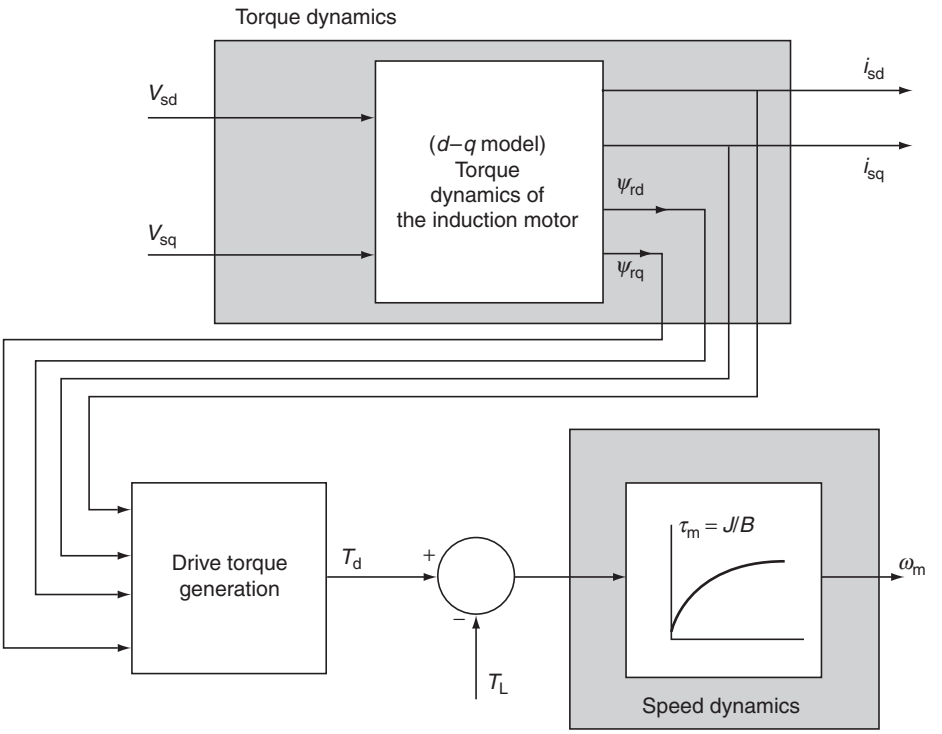


Figure 10.71 | Induction motor plant split as torque and speed dynamical systems.

q-controller) are designed to take care of the torque dynamics. The torque dynamics can be analyzed and the current controllers can be synthesized using only the torque model. As the mechanical time constant of the machine is generally very large compared to the electrical time constant of the induction motor, a reduced order model involving only the speed dynamics can be used to synthesize the speed controller by ignoring the current control loop dynamics.

Two digital current controllers, the *d*-axis and the *q*-axis controllers, can be synthesized simultaneously. In other words, the *d*-axis and the *q*-axis current control loops will be closed simultaneously. Later the speed controller can be synthesized using a reduced order model of the induction motor. In this way, output feedback can be used for synthesizing the current and speed controllers to obtain the controller gains. The torque model is a two-input, two-output system for the current controllers and a SISO system for the speed controller. The controller gains are obtained with the added advantage that classical intuition can still be brought forth in the choice of the dynamic control structure for the current- and the speed-controlled system.

In building the system model of the induction motor drive system with output feedback, one has to consider the dynamics of the compensator and other delays that are encountered in the control loop. Referring to Figure 10.71, the dynamics of the induction motor model that is transformed to the discrete domain and the dynamics of the *d*-axis and *q*-axis compensators form a part of the system formulation. The delay due to computation and the inverter delays will also have to be taken in account. Rather than exactly calculating the

computation and the inverter delays, one could include all delays in the control loop into one sample time. A sample time delay can be introduced in the *d*-axis and the *q*-axis loops to account for the various delays in the loops as shown in Figure 10.71. To facilitate modeling of the induction motor drive system to conform to the output feedback control structure of Figure 10.70, the inverters before the ZOH blocks have been introduced.

Since the field-oriented control is performed using a digital processor, the analog stator currents should be filtered to remove anti-aliasing effects after sampling. Also because the anti-alias filter is in the control path, the dynamics of the anti-alias filter should also be included in the system formulation. As the two-phase to three-phase and three-phase to two-phase transformations are algebraic operations, the plant that is considered is an equivalent two-phase induction motor that is described by Eq. (8.128a).

From the knowledge and experience with induction motor, one can arrive at a suitable compensator structure. The compensator shown in Figure 10.72 has an integrator and a proportional component (proportional–integral compensator). It should be noted that high dynamic performance is expected from the induction motor (i.e., the plant) when the arbitrary axis is oriented along one of the flux linkages of the motor. In the case of rotor field-oriented control, along with the use of the decoupling network the motor model reduces to a second-order system. This is because under rotor field orientation, the ψ_{rq} components vanish and the ψ_{rd} component is maintained constant. Each of the current control loops for i_{sd} and i_{sq} now see only a first-order plant. On the other hand, when the field is not oriented, then each current control loop sees a fourth-order plant.

It is important to note that the performance requirements are met only under the condition of rotor field orientation. But, if there exists some field disorientation, then the performance criteria like speed of

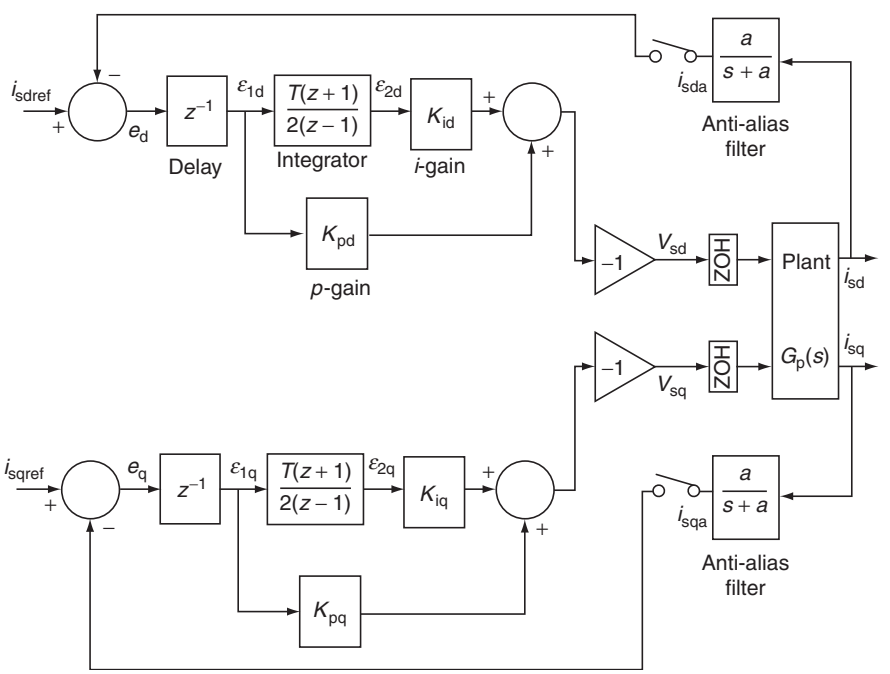


Figure 10.72 | Controller design block diagram.

response, overshoots, undershoots, etc. will not be met. However, the objective is to synthesize controllers wherein the stability and performance criteria are met when the arbitrary axis is oriented along the rotor field and at least stability criteria is met when the arbitrary axis is not oriented along the rotor field (wherein the plant is a fourth-order system). Therefore, it is important that the stability of the system is maintained even though there may be loss of performance due to field disorientation.

Plant Dynamics

The induction motor dynamics for torque or current control is given by Eq. (8.128a). Here, the state variables and the input variables (i_{sd}, i_{sq}, ψ_{rd}, ψ_{rq}, V_{sd} and V_{sq}) are re-defined to represent deviations in the neighborhood of an equilibrium point. The slip is not considered as an input but the variations in the slip of the induction motor are equivalently considered as being caused due to the uncertainty in the rotor time constant. Robustness to such uncertainties in the parameters of the induction motor will be treated in the next chapter. The plant G_p is described by matrices A_p, B_p, C_p and D_p which are given in the continuous domain by

$$A_p = \begin{bmatrix} -1/\tau_\sigma & \omega_m + \omega_{sl} & v_r / L_\sigma \tau_r & v_r \omega_m / L_\sigma \\ -(\omega_m + \omega_{sl}) & -1/\tau_\sigma & v_r \omega_m / L_\sigma & v_r / L_\sigma \tau_r \\ R_r v_r & 0 & -1/\tau_r & \omega_{sl} \\ 0 & R_r v_r & -\omega_{sl} & -1/\tau_r \end{bmatrix}$$

$$B_p = \begin{bmatrix} 1/L_\sigma & 0 \\ 0 & 1/L_\sigma \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$C_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$D_p = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{10.104}$$

where the plant state is given by

$$x_p = [i_{sd}, i_{sq}, \psi_{rd}, \psi_{rq}]^T$$

The input to the plant is given by

$$u_p = [V_{sd}, V_{sq}]^T$$

The output of the plant is given by

$$y_p = [i_{sd}, i_{sq}]^T$$

Anti-Alias Filter Dynamics and Augmentation

The plant $G_p(s)$ (i.e., the induction motor) is a low-pass filter. As long as the sampling frequency is selected at least twice as large as the plant cut-off frequency, the effects of aliasing will be small. However, if current sensors are used to measure the stator currents, then the sensors will introduce measurement noise. The high-frequency measurement noise may be aliased down to lower frequencies that are within the plant

bandwidth and thus have a detrimental effect on the system performance. To avoid this, a low-pass anti-aliasing filter of the form

$$H_a = \frac{a}{s+a} \tag{10.105}$$

is inserted after the measuring devices and before the samplers.

If the cut-off frequency of the anti-aliasing filter is not much higher than the plant cut-off frequency, then the filter will affect the closed-loop performance due to the finite attenuation and phase shift introduced to the measured signals. Therefore, the filter dynamics should be appended to the plant at the design stage itself so that the controller is designed taking it into account. As there are two measured outputs (i_{sd} and i_{sq}), there will be two anti-aliasing filters of the form indicated in Eq. (10.105), one for filtering i_{sd} and the other for filtering i_{sq}. Therefore, if G_a represents the anti-alias filter for both the axes, then the corresponding matrices, A_a, B_a, C_a and D_a, are given by

$$\begin{aligned} A_a &= \begin{bmatrix} -a & 0 \\ 0 & -a \end{bmatrix}; \quad B_a = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \\ C_a &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \qquad D_a = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{10.106}$$

where the anti-alias filter state is given by

$$x_a = [i_{sda}, i_{sqa}]^T$$

The input to the anti-alias filter is given by

$$u_a = [i_{sd}, i_{sq}]^T$$

The output of anti-alias filter is given by

$$y_a = [i_{sda}, i_{sqa}]^T$$

The induction motor plant G_p and the anti-alias filter G_a are in the continuous domain (s-domain) whereas the controllers are in the discrete domain (z-domain). The discrete and continuous domains are interfaced using ZOH and samplers as shown in Figure 10.72. The plant-anti-alias filter augmented system G_{pa} is described by the matrices A_{pa}, B_{pa}, C_{pa} and D_{pa} which are given by

$$A_{pa} = \begin{bmatrix} a_{pa11} & a_{pa12} \\ a_{pa21} & a_{pa22} \end{bmatrix} \qquad (6 \times 6 \text{ matrix}) \tag{10.107a}$$

where

$$a_{pa11} = A_p \qquad (4 \times 4 \text{ matrix})$$

$$a_{pa12} = 0 \qquad (4 \times 2 \text{ matrix})$$

$$a_{pa21} = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \end{bmatrix}$$

$$a_{\text{pa22}} = A_{\text{a}} \qquad (2 \times 2 \text{ matrix})$$

$$B_{\text{pa}} = \begin{bmatrix} b_{\text{pa11}} \\ b_{\text{pa21}} \end{bmatrix} \qquad (6 \times 2 \text{ matrix}) \tag{10.107b}$$

where

$$b_{\text{pa11}} = B_{\text{p}} \qquad (4 \times 2 \text{ matrix})$$

$$b_{\text{pa21}} = 0 \qquad (2 \times 2 \text{ matrix})$$

The state of G_{pa} is given by

$$x_{\text{pa}} = [x_{\text{p}}, x_{\text{a}}]^{\text{T}}$$

The input to G_{pa} is given by

$$u_{\text{pa}} = u_{\text{p}}$$

The output of G_{pa} is given by

$$y_{\text{pa}} = y_{\text{a}}$$

The augmented plant $G_{\text{pa}}(s)$ is transformed to the discrete domain using the ZOH equivalent transformation to obtain $G_{\text{pa}}(z)$.

Controller Dynamics and Augmentation

Referring to Figure 10.72, the unit delay term is used to include all the computational and loop delays that may occur. This unit delay is shifted before the proportional–integral control structure to make the controller formulation easier without loss of generality.

Considering first the delay dynamics, one obtains

$$\begin{bmatrix} \varepsilon_{\text{1d}} \\ \varepsilon_{\text{1q}} \end{bmatrix}_{k+1} = \begin{bmatrix} -i_{\text{sda}} \\ -i_{\text{sqa}} \end{bmatrix}_{k} + \begin{bmatrix} i_{\text{sdref}} \\ i_{\text{sqref}} \end{bmatrix}_{k} \tag{10.108}$$

Considering now the proportional–integral dynamics, one obtains

$$\begin{bmatrix} \varepsilon_{\text{2d}} \\ \varepsilon_{\text{2q}} \end{bmatrix}_{k+1} = \begin{bmatrix} (T/2)\varepsilon_{\text{1d}}\varepsilon_{\text{2d}} \\ (T/2)\varepsilon_{\text{1q}}\varepsilon_{\text{2q}} \end{bmatrix}_{k} + \begin{bmatrix} (T/2)\varepsilon_{\text{1d}} \\ (T/2)\varepsilon_{\text{1q}} \end{bmatrix}_{k+1} \tag{10.109}$$

where T is the sampling time of the analog signals. Substituting Eq. (10.108) in Eq. (10.109), the controller dynamics is given by

$$\begin{bmatrix} \varepsilon_{\text{1d}} \\ \varepsilon_{\text{1q}} \\ \varepsilon_{\text{2d}} \\ \varepsilon_{\text{2q}} \end{bmatrix}_{k+1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{T}{2} & 0 & 1 & 0 \\ 0 & \frac{T}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{\text{1d}} \\ \varepsilon_{\text{1q}} \\ \varepsilon_{\text{2d}} \\ \varepsilon_{\text{2q}} \end{bmatrix}_{k} + \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ -\frac{T}{2} & 0 & \frac{T}{2} & 0 \\ 0 & -\frac{T}{2} & 0 & \frac{T}{2} \end{bmatrix} \begin{bmatrix} i_{\text{sda}} \\ i_{\text{sqa}} \\ i_{\text{sdref}} \\ i_{\text{sqref}} \end{bmatrix}_{k} \tag{10.110}$$

where

$$A_c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{T}{2} & 0 & 1 & 0 \\ 0 & \frac{T}{2} & 0 & 1 \end{bmatrix}$$

$$B_c = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ -\frac{T}{2} & 0 & \frac{T}{2} & 0 \\ 0 & -\frac{T}{2} & 0 & \frac{T}{2} \end{bmatrix}$$

The state of G_c is given by

$$x_c = [\varepsilon_{1d}, \varepsilon_{1q}, \varepsilon_{2d}, \varepsilon_{2q}]^T$$

The input to G_c is given by

$$u_c = [i_{sda}, i_{sqa}, i_{sdref}, i_{sqref}]^T$$

The output of G_c is given by

$$y_c = x_c$$

Augmenting the controller G_c to the plant-filter dynamics G_{pa}, the total system dynamics G is obtained. If the total system (plant-filter and controller) is described by the matrices A, B, C, E and F as described in Eqs. (10.101a), (10.101b), (10.101c), one obtains

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad (10 \times 10 \text{ matrix}) \qquad (10.111)$$

where

$$a_{11} = A_{pa} \qquad (6 \times 6 \text{ matrix})$$

$$a_{12} = 0 \qquad (6 \times 4 \text{ matrix})$$

$$a_{21} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & \frac{-T}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{-T}{2} \end{bmatrix}$$

$$a_{22} = A_c \qquad (4 \times 4 \text{ matrix})$$

$$B = \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \qquad (10 \times 2 \text{ matrix}) \tag{10.112}$$

where

$$b_{11} = B_{\text{pa}} \qquad (6 \times 2 \text{ matrix})$$

$$b_{21} = 0 \qquad (4 \times 2 \text{ matrix})$$

$$E = \begin{bmatrix} e_{11} \\ e_{21} \end{bmatrix} \qquad (10 \times 2 \text{ matrix}) \tag{10.113}$$

where

$$e_{11} = 0 \qquad (6 \times 2 \text{ matrix})$$

$$e_{21} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \dfrac{T}{2} & 0 \\ 0 & \dfrac{T}{2} \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{10.114}$$

$$F = 0 \qquad (4 \times 2 \text{ matrix}) \tag{10.115}$$

For the augmented total system shown in Figure 10.73, the state of G is given by

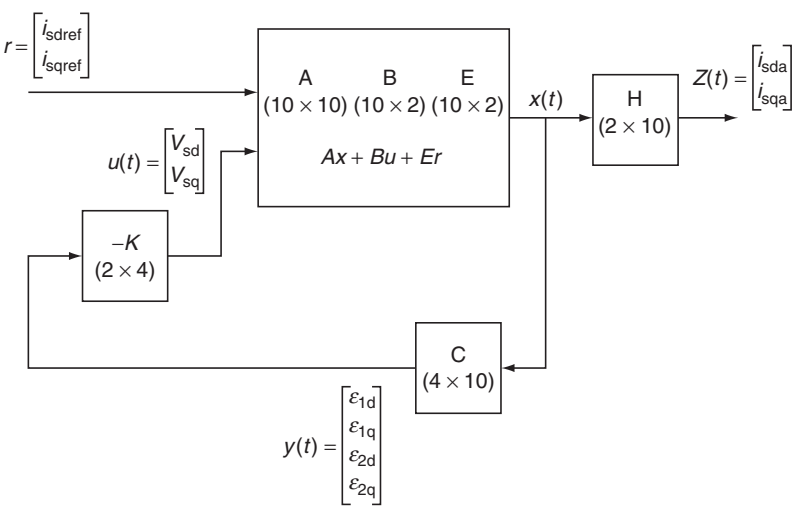


Figure 10.73 Augmented system of the current control system for the induction motor with output feedback.

$$x = [x_{pa}, x_c]^T \text{ (totally there are 10 state variables)}$$

The input to G is given by

$$u = u_{pa} = u_p$$

The reference input of G is given by

$$r = [i_{sdref}, i_{sqref}]^T$$

The output of G is given by

$$y = y_c$$

The performance output for the system is

$$z = [i_{sda}, i_{sqa}]^T = Hx$$

and the performance output matrix H is given by

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Referring to Figure 10.72, the control input to the plant-filter u_{pa} is given by

$$u_{pa} = \begin{bmatrix} V_{sd} \\ V_{sq} \end{bmatrix} = -\begin{bmatrix} K_{pd} & 0 & K_{id} & 0 \\ 0 & K_{pq} & 0 & K_{iq} \end{bmatrix} \begin{bmatrix} \varepsilon_{1d} \\ \varepsilon_{1q} \\ \varepsilon_{2d} \\ \varepsilon_{2q} \end{bmatrix} \tag{10.116}$$

which is of the accepted control law $u = -Ky$ where

$$K = \begin{bmatrix} K_{pd} & 0 & K_{id} & 0 \\ 0 & K_{pq} & 0 & K_{iq} \end{bmatrix}$$

From Eqs. (10.111)–(10.116), the induction motor is now formulated as an output feedback structure as given in Eq. (10.101). The dynamics of the system is determined by $A - BKC$ which is used to design the value of K as discussed in the previous section.

CONCLUDING REMARKS

Controller design is one of the aspects of a closed-loop system that is most essential for the system performance but least understood from the implementation perspective. This chapter endeavors to give controller design insights through illustrative power electronic system examples.

The classical controller design methods are still popular but for smaller systems. The Bode diagram especially is very popular for stable non-minimum phase systems. However, the Bode diagram has its limitations as discussed in this chapter. The root locus method is definitely more complete in the sense that it can handle any type of system as long as it is linear and time invariant.

The state space methods are very powerful. The only uncertainty is in modeling of the physical system.

The quality of the performance is directly related to the accuracy of the model of the physical system. Once the modeling process is successfully accomplished, the state space methods can give closed-form solutions with guaranteed stability. Compare and contrast this with the classical methods wherein there are a number of uncertainties in the controller design process. The modeling of the system, the choice of the controller pole–zero structure and the choice of the controller gain are open aspects leading to many trials before obtaining a successful controller.

The estimator design is another important aspect of the state space approach. This can bring in greater insight about the system and also reduce the number of physical sensors that one needs to use. Consequently, the cost and the reliability of the overall system will improve. The controller design is also an art. Experience and intuition can lead to quick and better controllers especially using the classical methods or the output feedback methods. The next section gives some tutorial exercises on the three primary DC–DC converters. These should be studied in the light of the discussions in this chapter, using a computer-aided programming environment for faster iteration and removing the drudgery of manual computation and fatigue.

TUTORIAL EXERCISES

The following tutorial exercises can be studied in either MATLAB or SciLAB environment.

1. Implement a PID module in a simulation environment with anti-windup feature.

2. The small-signal dynamic transfer function for the buck converter is given below. The control input is the duty cycle input pin and the controlled output is the output voltage V_o. The main performance objective is to attain zero steady-state error. Choose the sampling frequency to be same as the switching frequency.

$$\frac{\hat{v}_o}{\hat{d}}(s) = \frac{V_i}{\left(1 + s\frac{L}{R} + s^2 LC\right)}$$

where V_i = 10 V; L = 10 mH; C = 100 μF; R = 10 Ω and switching frequency is 10 kHz. V_i is the input source voltage, d is the operating duty cycle, R is the load resistance and V_o is the output voltage and C is the output capacitor. Design digital controller for the above system, using both the root locus approach and the full-state feedback approach. (Note: capacitor voltage is a state variable and the inductor current is a state variable.)

3. The small-signal dynamic transfer function for the boost converter is given below. The control input is the duty cycle input pin and the controlled output is the output voltage V_o. The main performance objective is to attain zero steady-state error. Choose the sampling frequency to be same as the switching frequency.

$$\frac{\hat{v}_o}{\hat{d}}(s) = \frac{V_i}{(1-D)^2} \frac{\left(1 - s\frac{L}{R(1-D)^2}\right)}{\left(1 + s\frac{L}{R(1-D)^2} + s^2\frac{LC}{(1-D)^2}\right)}$$

where V_i = 10 V, L = 1 mH, C = 100 μF, R = 20 Ω; D = 0.4 and switching frequency is 10 kHz. V_i is the input source voltage, D is the nominal operating duty cycle, R is the load resistance, V_o is the output voltage and C is the output capacitor.

(a) Design digital controller for the above system using full-state feedback approach with and without full-order estimator and simulate the complete system.

(b) For a reference input tracking problem, design K, N_x, N_u and G (full-order prediction observer) and simulate the complete system.

4. The small-signal dynamic transfer function for the buck–boost converter is given below. The control input is the duty cycle input pin and the controlled output is the output voltage V_o. The main performance objective is to attain zero steady-state error. Choose the sampling frequency to be same as the switching frequency.

$$\frac{\hat{v}_o}{\hat{d}}(s) = \frac{-V_i}{(1-D)^2} \frac{\left(1 - s\dfrac{DL}{R(1-D)^2}\right)}{\left(1 + s\dfrac{L}{R(1-D)^2} + s^2\dfrac{LC}{(1-D)^2}\right)}$$

where $V_i = 10$ V, $L = 1$ mH, $C = 100$ uF, $R = 20\ \Omega$, $D = 0.4$ and switching frequency is 10 kHz.

V_i is the input source voltage, D is the nominal operating duty cycle, R is the load resistance, V_o is the output voltage and C is the output capacitor.

(a) Design digital controller for the above system using full-state feedback approach with and without full-order estimator and simulate the complete system.

(b) For a reference input tracking problem, design K, N_x, N_u and G (full-order prediction observer) and simulate the complete system.

FILL IN THE BLANKS

1. The controller design addresses the three performance criteria, _____, _____ and _____.

2. By Bode diagram approach, it is important to maintain the slope of the amplitude–log frequency curve in the vicinity of the gain crossover frequency at a slope of _____.

3. Bode diagram method should not be applied for _____ phase systems.

4. In Bode diagram the magnitude is given as _____ and the phase angle is _____.

5. In Bode diagram, the two parameters used for the controller design are _____ and _____.

6. Both gain and phase margins are measures of _____.

7. The steady-state error is addressed by the _____ of the PID controller.

8. The proportional and derivative parts of the PID controller address the _____ criteria.

9. Noise is dominant at _____ frequencies.

10. The order of a practical PID controller is _____.

11. The root locus method is a graphical technique for determining the _____ poles of a system as a function of the _____.

12. The transfer function assumes _____ conditions.

13. The transfer function is applicable only for linear time-invariant _____ input single _____ systems.

14. The state space approach is applicable for _____ input _____ output systems.

15. A system is controllable if there exists a sequence of _____ such that it transfers the system from any _____ state to any final state in _____ time.

16. A system is said to be _____ provided that any initial state can be calculated from finite number of _____.
17. The controller for a full-state feedback system is a matrix of _____.
18. In full-state feedback control, the closed-loop poles of the system can be placed _____ within the unit circle.
19. The controller design and the estimator gain design can be done _____ and _____ of each other.
20. To ensure that the power calculated in the stationary axis is the same as the power calculated in the rotating reference axis, a factor of _____ is used as the power balance factor.
21. *V*/*f* control of induction motor falls in the class of _____ control.
22. In stator voltage control if the stator voltage is reduced, the flux _____.
23. The region beyond base frequency is called the _____ region.
24. To compensate for the _____ drop, a small boost to the stator terminal voltage is provided at low frequencies.
25. The torque producing and the flux producing components of the stator current are_____ in field-oriented control of induction motor.
26. The output feedback method considers only _____ state feedback instead of full-state feedback.
27. The output feedback control law can be used to design compensators with any desired _____ structure, thereby regaining much of the intuition of classical controls design.
28. In output feedback the controller gain is determined so that both the closed-loop _____ and _____ are placed within the unit circle in the *z*-plane.

DESCRIPTIVE QUESTIONS

1. What are the classical and modern methods of controller design?
2. Write short notes on: (a) Bode's theorem I and (b) Bode's theorem II.
3. What is a non-minimum phase system? Give an example.
4. Discuss Bode diagram construction.
5. Write short notes on: (a) Gain margin, (b) Phase margin.
6. Discuss the design steps in the controller design with Bode diagram.
7. Discuss the PID controller.
8. Explain the step response of a closed-loop system that has a PID as its controller.
9. What is the transfer function of the PID controller?
10. How are the constants K_i, K_p and K_d for a PID controller selected?
11. Write short notes on: (a) Integrator windup, (b) anti-windup.
12. What is a root locus? Explain.
13. Discuss the design steps for designing a controller using the root locus technique.
14. Write short notes on: (a) continuous state equation, (b) discrete state equation.
15. Write short notes on: (a) controllability and (b) observability.
16. What is full-state feedback?

17. Distinguish between regulator problem and tracking problem?

18. What is the regulator control law? Explain.

19. Discuss regulator design by pole placement.

20. What is an estimator? Is there a difference between an estimator and an observer?

21. Discuss the full-order prediction estimator.

22. What is the estimator state equation for a full-order prediction estimator?

23. Discuss the design steps to obtain the prediction estimator gain matrix.

24. How can the same Ackermann's formula be used for obtaining the prediction estimator gain matrix?

25. What is separation principle? Explain.

26. What is the dynamics of the controller and estimator together?

27. Discuss the full-order current estimator.

28. What is the estimator state equation for a full-order current estimator?

29. Explain the concept of time update and measurement update in current estimator.

30. Discuss the design steps to obtain the current estimator gain matrix.

31. How can the same Ackermann's formula be used for obtaining the current estimator gain matrix?

32. Why is reduced order estimator needed? What are its benefits as compared to a full-order estimator?

33. Explain reduced order prediction estimator?

34. Explain reduced order current estimator?

35. How can the same Ackermann's formula be used for obtaining the reduced order estimator gain matrix?

36. Discuss the tracker problem? Is the controller design same as that of the regulator problem?

37. How are N_x and N_u matrices designed for a specific system?

38. How is the output voltage fed back to the controller side of the DC–DC converter with opto-isolation? How is the feedback opto-coupler's temperature drift compensated?

39. What are the different ways of regulating the outputs of multiple-output DC–DC converters?

40. Write short notes on: (a) coupled inductor, (b) magnetic amplifier.

41. What are the drawbacks of the coupled inductor approach to multiple-output regulation?

42. In current control of forward converter, why should the duty cycle be limited to 0.5?

43. What is the principle behind slope compensation in current programmed converters?

44. Explain the operation of the boost converter as a unity power factor converter?

45. Discuss the reference current generation for single-phase unidirectional unity power factor converter?

46. How is the inverter voltage reference generated for single-phase front-end converter in order to improve the power factor?

47. Discuss the single-phase front-end converter operation and how it can be extended to three-phase systems?

48. Explain the vector control of a three-phase front-end converter for unity power factor at the grid.

49. What is power balance factor in vector-controlled systems?

50. Discuss the steady-state equivalent circuit of an induction motor.

51. Why should a stator voltage boost be provided at low applied frequencies?

52. Discuss the frequency versus applied voltage characteristics that is required for *V*/*f* control? Compare with the frequency versus back emf characteristics.

53. Distinguish between open-loop *V*/*f* and closed-loop *V*/*f* operation.

54. What are the main conceptual differences between *V*/*f* control and vector control of induction motor?

55. What is achieved by aligning the synchronously rotating reference frame along the rotor flux axis?

56. Write short notes on: (a) direct field-oriented control, (b) indirect field-oriented control.

57. Discuss the block schematic of the direct and indirect vector control of induction motor.

58. What is decoupling network in vector control of induction motor? Discuss with respect to the rotor field-oriented control.

59. What is the control law for output feedback?

60. What are the benefits and drawbacks of output feedback vis-à-vis the full-state feedback?

61. How can the Ackermann's formula be used to determine the output feedback controller gain matrix?

PROBLEMS

1. A unity feedback control system has an open-loop transfer function given as

$$G(s) = \frac{10}{s(1+0.2s)(1+0.5s)}$$

Plot the Bode diagram. Find out the gain cross-over frequency. What is the gain margin and phase margin?

2. A unity feedback control system has an open-loop transfer function given as

$$G(z) = \frac{z(z+1)(z+0.2)}{(z+0.3)(z^2-0.5z+0.5)}$$

Plot the Bode diagram. Find out the gain cross-over frequency. What is the gain margin and phase margin?

3. A unity feedback control system has an open-loop transfer function given as

$$G(s) = \frac{10}{s(1+0.2s)(1+0.5s)}$$

Plot the root loci. Is the root loci entirely within the stable region of the *s*-plane? At what value of the controller gain, does the system become unstable?

4. A unity feedback control system has an open-loop transfer function given as

$$G(z) = \frac{z(z+1)(z+0.2)}{(z+0.3)(z^2-0.5z+0.5)}$$

Plot the root loci. Is the root loci entirely within the stable region of the *s*-plane? At what value of the controller gain, does the system become unstable?

5. Derive Ackermann's formula for controller gain calculation.

6. Derive the error dynamics for a full-order prediction estimator?

7. Derive the error dynamics for a full-order current estimator?

8. Suggest methods other than opto-isolation, to feedback the isolated output voltage to the

controller side of the DC–DC converter without sacrificing galvanic isolation.

9. Show that the flux in the induction machine is constant if the ratio of back emf to the applied frequency is maintained constant.

10. Deduce the rotor flux estimate model for rotor field-oriented control.

11. Deduce the stator flux estimate model for rotor field-oriented control.

12. A plant transfer function in the continuous domain is given as

$$G_p(s) = \frac{K_g(s+b)\omega_r^2}{s^2(s^2 + 2\zeta\omega_r s + \omega_r^2)}$$

where the sampling rate is 0.1 ms, $b = 0.0001$, $\zeta = 0.005$, $\omega_r = 500$, $K_g = 0.5$.

Plot the Bode diagram and find out the gain and phase margins. Design a controller for a phase margin of atleast 60° and a gain margin of atleast 6 dB.

13. For Problem 12, plot the root loci and design a controller using the root locus technique.

14. Consider servomotor transfer function given as $10/[s(s+1)]$ wherein the output is the position of the shaft. The state variables of the system can be considered to be the shaft position and speed. Using a sampling rate of 0.1 s, convert the continuous transfer function to discrete domain using zero-order hold. From the discrete transfer function obtain the difference equations giving the discrete state equations of the system.

15. For Problem 14, design the following:

a. Full-state feedback controller

b. Dead beat controller

c. Full-order prediction estimator

d. Full-order current estimator

e. Reduced order prediction estimator

f. Reduced order current estimator

g. N_x and N_u matrices for the tracking problem

16. Using the root locus technique, design controllers for the following plants for zero steady-state error and fastest response with zero overshoot:

a. $\dfrac{1}{z(z-1)}$

b. $\dfrac{1}{z(z-0.8)}$

c. $\dfrac{(z-4)}{(z-1)}$

d. $\dfrac{1}{(z-1)^2}$

e. $\dfrac{z^2}{(z-1)^2}$

ANSWERS

Fill in the Blanks

1. stability; steady-state accuracy; speed of response
2. −20 dB/decade
3. non-minimum
4. $20 \cdot \log_{10} G(j\omega)$; $\angle G(j\omega)$
5. gain margin; phase margin
6. relative stability
7. integral or I-part
8. speed of response
9. higher
10. two
11. closed loop; gain
12. zero initial
13. single; output
14. multiple; multiple

15. inputs; initial; finite
16. observable; measurements
17. gains
18. anywhere
19. separately; independent
20. 2/3
21. scalar
22. decreases
23. field weakening
24. stator resistance
25. decoupled
26. measurable
27. dynamical
28. poles; zeros

CHAPTER 11

Optimal and Robust Controller Design

Learning Objectives

After reading this chapter, you will be able to:

- understand and apply optimal controller concepts.
- apply least squares and weighted least squares algorithm.
- apply recursive algorithms such as least mean square and Kalman filter algorithms.
- design optimal controller and estimator gains.
- understand robust controller basics.

Based on the discussions in Chapter 10, it may be noted that there are significant advantages in using the state space approach for the design of controllers. They may broadly be classified as full-state feedback and output feedback. In full-state feedback all the state variables of the system are required to be fed back to the controller to take appropriate control action. As all the state variables are fed back, the entire information about the system is available in these state variables. Hence only scalar gains are needed to appropriately scale these state variables to obtain the necessary control performance. Thus in full-state feedback approach the introduction of the controller does not increase the overall order of the system with respect to the uncontrolled open-loop system.

On the other hand, in output feedback, only the measurable states are fed back to the controller. Hence the controller does not have the entire information about the state of the system as in the case of a full-state feedback. As a consequence, dynamical controllers are needed to introduce additional degrees of freedom in the synthesis of the controllers to achieve the stated performance. Thus, the overall system will be of a higher order with respect to the order of the uncontrolled open-loop system. This is much like the single-input single-output (SISO) case wherein only one state variable is fed back to the controller with only partial information about the state of the system. The output feedback brings in the intuitive insights and experience that one uses in the classical SISO controller designs.

Further, in the case of the full-state feedback approach, many a times all the state variables are not available to be fed back due to either inaccessibility or the sensors may be too expensive. To overcome this problem, the unavailable states are estimated by using estimators or observers that make use of the information in the measurable states to estimate the state variable values of the unavailable states. This way the controller is fed with all the state variables for the purposes of control. Therefore, in the design of control systems based on the full-state feedback approach, two major blocks need to be synthesized: (a) controller; (b) estimator or observer. This implies that the controller gain K and the estimator gain G need to be designed. The previous

chapter discussed various methods of designing the controller gain K and the estimator gain G to meet the primary performance criteria (SSS), that is, the stability, steady-state accuracy and speed of response.

Suppose one would like to design K and G to meet the primary SSS performance criteria under some constraints like minimum input energy or minimum state energy. Such controllers and estimators that have been designed to meet the performance criteria under some constraints are called optimal controller and optimal estimator, respectively. Kalman gain is one such optimal controller that achieves the control performance criteria under the constraints of minimum input and state energies in the presence of system noise. Kalman filter is an optimal estimator that estimates the unavailable states of a system in the presence of noise. The optimal controllers and the optimal estimators operate on the linear quadratic principles which will be the focus of discussion in this chapter. The topic will be developed first by starting with the estimation and later leading to the discussion on the optimal control aspects.

11.1 Least Squares Principle

Consider a system that is represented as a state space model given as

$$\dot{x} = Ax + Bu \tag{11.1}$$

$$y = Cx + Du \tag{11.2}$$

Referring to the output equation [Eq. (11.2)], in most practical situations, there is no direct feed through from the input to the output. Hence D is a null matrix. Thus in majority of the situations,

$$y = Cx \tag{11.3}$$

where y is the output that is measured, x is the state vector and C gives the relationship between the output and the state vector. Let y be the output value that is given by Eq. (11.3) and y_{act} represent the actual value of the output. As the mathematical representation of the system [Eqs. (11.1)–(11.3)] will never be the exact system representation, in practice there will always be an error e between the model value y and the actual value y_{act}. Thus,

$$Cx = y \neq y_{act}$$

The error is defined as

$$\begin{aligned} e &= y_{act} - y \\ &= y_{act} - Cx \end{aligned}$$

If $y_{act} - Cx = 0$ then the solution x is trivial and is given by $C^{-1}y_{act}$. However, in practice, $y_{act} - Cx$ is never equal to zero. There will always exist an error between the actual or measured values and the model values as given by Cx. In such a situation, there is no trivial solution. One will have to find the value of x that will minimize the error $y_{act} - Cx$. Such a solution for x is the least square solution. The following discussion will evolve the least square solution for $y_{act} - Cx) = e$.

Consider now a series of measurements of the actual value of the output, that is, y_{act1}, y_{act2}, …. as shown in Figure11.1. The equation $y = Cx$ gives the best estimate of the actual measured points of the output as indicated in Figure 11.1. The central function of an estimator is to provide the value of x which will minimize the errors. If the objective is to minimize the error, one should not consider the sum of the errors as the function to be minimized because the positive and negative errors may cancel each other and give a wrong notion about the goodness of the fit. Alternately one may consider minimizing the sum of the error squares (e^2).

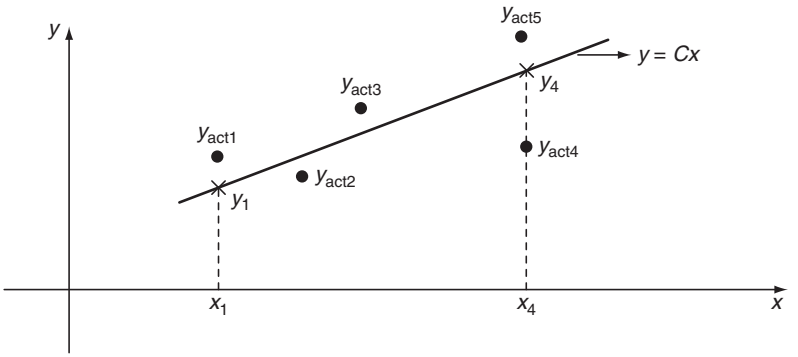


Figure 11.1 | Actual values and the model values.

Referring to Figure 11.1, one has

$$\begin{aligned} e_1 &= y_{\text{act}1} - y_1 \\ e_2 &= y_{\text{act}2} - y_2 \\ &\vdots \\ e_n &= y_{\text{act}n} - y_n \end{aligned} \tag{11.4}$$

Let the function to be minimized be

$$f(x) = \sum_{i=1}^{n} e_i^2$$

This can be written elegantly in vector form as

$$f(x) = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{e}^{\mathrm{T}}\mathbf{e} \tag{11.5}$$

Equation (11.5) is the function that is to be minimized by choosing proper values of x. The value of x that gives minimum value of $f(x)$ is the best estimate of the state vector. The solution x of Eq. (11.3) is that value of x which will provide the minimum value of $f(x)$ and is called the least squares solution. Before obtaining the solution for Eq. (11.3) based on the constraint that the solution should minimize Eq. (11.5), it is appropriate at this point to slightly deviate to understand the quadratics in matrix notation called the quadratic forms.

11.2 Quadratic Forms

A quadratic form in two variables x_1 and x_2 is an expression that can be written as

$$f = ax_1^2 + 2bx_1x_2 + cx_2^2$$

where f is called the quadratic function. The function f can be written in matrix form as

$$f = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{11.6}$$

$$= \mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x}$$

Equation (11.6) is a quadratic of two dimensions. In general, a quadratic form in "n" variables $x_1, x_2, \ldots, x_n$ is defined to be an expression that can be written as

$$\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \mathbf{C} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} \tag{11.7}$$

where $\mathbf{C}$ is an $n \times n$ matrix. Equation (11.7) is called a quadratic of n dimensions. Consider the function

$$f(x_1, x_2) = ax_1^2 + 2\,bx_1x_2 + cx_2^2$$

where a, b and c are non-zero. It can be seen that f is zero only when $x_1 = x_2 = 0$, that is, when $\mathbf{x} = \mathbf{0}$, a null matrix. Now the question is when $\mathbf{x} \neq \mathbf{0}$ then is $f > 0$ or is $f < 0$? For the moment, consider that $\mathbf{C}$ in Eq. (11.6) is an identity matrix. Then,

$$f = \mathbf{x}^{\mathrm{T}}\mathbf{x}$$

This results in f being a sum of all square terms which cannot be negative whatever be the non-zero value of $\mathbf{x}$. Therefore, if $\mathbf{C}$ is identity matrix, then f is always positive (i.e., $f > 0$). This implies that f being positive or negative is solely determined by the character of the $\mathbf{C}$ matrix which contains the coefficients of the quadratic function f. Thus the following cases arise:

1. If f is positive ($f > 0$) for all non-zero values of $\mathbf{x}$, implying $\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} > 0$, such a matrix $\mathbf{C}$ is called positive definite.
2. If f is negative ($f < 0$) for all non-zero values of $\mathbf{x}$, implying $\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} < 0$, such a matrix $\mathbf{C}$ is called negative definite.
3. If f is zero or positive ($f \geq 0$) for all non-zero values of $\mathbf{x}$, implying $\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} \geq 0$, such a matrix $\mathbf{C}$ is called positive semi-definite.
4. If f is zero or negative ($f \leq 0$) for all non-zero values of $\mathbf{x}$, implying $\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} \leq 0$, such a matrix $\mathbf{C}$ is called negative semi-definite.
5. If f is zero or positive or negative for all non-zero values of $\mathbf{x}$, implying $\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x}$ can either be zero, positive or negative, such a matrix $\mathbf{C}$ is called indefinite.

Figure 11.2 illustrates the concept of positive definiteness, negative definiteness and indefiniteness for a two-dimensional quadratic form, that is, the case of a function f of two variables x_1 and x_2. In Figure 11.2, x_1 and x_2 are the two orthogonal axes in a plane and f is the axis that is orthogonal to the plane formed by x_1, x_2.

Figure 11.2(a) indicates a cup-shaped paraboloid wherein f is always positive whatever be the non-zero value of $\mathbf{x} = [x_1, x_2]$. This means that $\mathbf{C}$ is positive definite. It also indicates that at $\mathbf{x} = [0, 0]$, f is zero implying that the minimum of the function f in this case is at $\mathbf{x} = \mathbf{0}$. Figure 11.2(b) indicates an inverted paraboloid wherein f is always negative whatever be the non-zero value of $\mathbf{x} = [x_1, x_2]$. This means that $\mathbf{C}$ is negative definite. It also shows that at $\mathbf{x} = [0, 0]$, f is zero implying that the maximum of the function f in this case

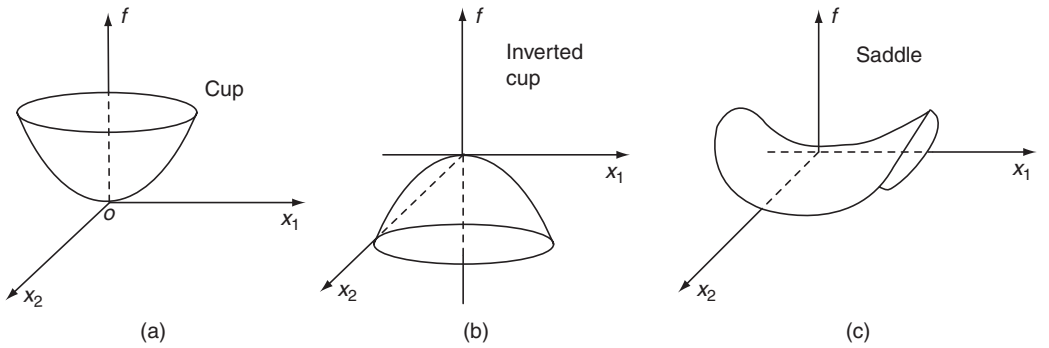


Figure 11.2 Two-dimensional quadratic form: (a) Positive definiteness; (b) negative definiteness; (c) indefiniteness.

is at $\mathbf{x} = \mathbf{0}$. In the case shown in Figure 11.2(c), the function f is shaped like a saddle. Here f takes both positive and negative values for non-zero values of $\mathbf{x}$. The function f is zero not only at $\mathbf{x} = \mathbf{0}$ but also at other non-zero values of $\mathbf{x}$. In such cases, the matrix $\mathbf{C}$ is called indefinite. There is neither a unique minimum nor a maximum. In general,

1. If $\mathbf{C}$ is positive definite, there always exists a minimum.
2. If $\mathbf{C}$ is negative definite, there always exists a maximum.

Now the question arises as to what is the practical significance of the quadratic forms? In physical systems, the energy is a quadratic form which is proportional to square of voltage or square of current or square of velocity or square of angular speed, etc. In general, the square of the effort or the square of the flow in a system represents the energy which is a quadratic form. The error square that was discussed earlier is also a quadratic form. Many situations arise in practice where one needs to minimize either the square of the error or the energy. Therefore, understanding the quadratic forms and the various definiteness of $\mathbf{C}$ is essential to the process of optimization.

11.3 Minimum Energy Principle

All physical systems will reach a stable state, also called the equilibrium state, where the system will be at its minimum energy state. As examples, a stick held vertically, will ultimately try to reach its minimum energy state which is the horizontal position; a child sliding down a slide will reach the minimum energy state on having reached the bottom; a capacitor connected to a load will discharge and reach the zero energy state; and so on. Nature sees to it that maximum amount of energy is removed from the system so that it may reach the equilibrium or the stable state.

Consider an *RC* circuit as shown in Figure 11.3. On closing the switch S the circuit will tend to reach an equilibrium state. A current i will flow in the circuit which will charge up the capacitor towards V_g. As the capacitor exponentially charges up to V_g, the current i will exponentially die down to zero. In the process of charging up the capacitor so that the circuit reaches an equilibrium state, nature utilizes some energy which is dissipated in the resistor R. Maximum energy will be removed from this circuit to reach equilibrium or minimum energy state. This is true in the case of any physical system. This is called the *minimum energy principle*.

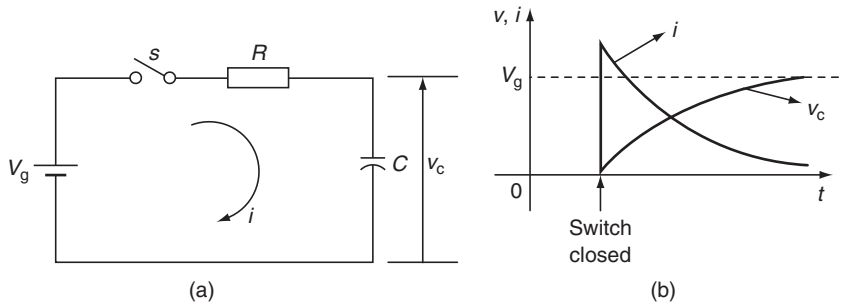


Figure 11.3 | (a) *RC* circuit; (b) current and voltage waveforms of capacitor.

Referring to the *RC* circuit of Figure 11.3, on closing the switch S, by applying the Kirchhoff's voltage law (KVL) around the loop, one has

$$V_g = Ri + v_c$$

Multiplying the above equation by i and integrating, one obtains the following energy equation:

$$\int V_g i\,dt = \underbrace{\int Ri^2\,dt}_{\text{Energy lost}} + \int v_c i\,dt \tag{11.8}$$

Equation (11.8) gives the energy equation for the *RC* circuit of Figure 11.3. The energy that is dissipated is the energy that is required by nature to establish equilibrium in the circuit. Let the energy lost in the resistor be denoted by E_R. Then Eq. (11.8) can be re-written as

$$E_R = \int V_g i\,dt - \int v_c i\,dt \tag{11.9}$$

From circuit theory, the current i through the *RC* circuit is given as

$$i = C\frac{dv_c}{dt}$$

Substituting for i in Eq. (11.9), one obtains

$$E_R = \int V_g C\frac{dv_c}{dt}\,dt - \int v_c C\frac{dv_c}{dt}\,dt$$

which can be simplified as

$$E_R = V_g C v_c - \frac{1}{2}Cv_c^2 \tag{11.10}$$

It is evident that Eq. (11.10) is a quadratic and the energy E_R is a parabola with respect to v_c. As nature prefers to bring the circuit to equilibrium or minimum energy state, maximum amount of energy is removed from the circuit by way of loss in R. One should find the maximum of the parabola of Eq. (11.10). Differentiating E_R with respect to v_c and equating to zero, one obtains the condition for maximum value of E_R. This is given by

$$\frac{dE_R}{dv_c} = V_g C - Cv_c = 0$$

$$v_c = V_g$$

Applying the above condition to the loss energy equation [Eq. (11.10)], one obtains

$$E_{\mathrm{R}} = CV_{\mathrm{g}}^2 - \frac{1}{2}CV_{\mathrm{g}}^2 = \frac{1}{2}CV_{\mathrm{g}}^2$$

This implies that whatever be the value of the resistor R, nature needs to remove the maximum amount of energy equivalent to $(1/2)CV_{\mathrm{g}}^2$ through the resistor to ensure that the circuit reaches equilibrium or the minimum energy state. The energy equation [Eq. (11.10)] can be generalized for any physical system. In general if E_{N} is the energy of any system then E_{N} has a minimum and the energy removed E_{R} has a maximum. The generalized energy equation is given as

$$E_{\mathrm{N}}(x) = \frac{1}{2}Ax^2 - bx$$

$$\frac{\mathrm{d}E_{\mathrm{N}}}{\mathrm{d}x} = Ax - b \tag{11.11}$$

where x is any state variable and A and b are parameters of the system. Here again the derivative is zero at $x = b/A$. If A is positive, the parabola opens towards the positive axis and this equilibrium point would be stable. If A is negative, the parabola opens towards the negative axis and this point $x = b/A$ would be a maximum and would not therefore be a stable point.

In the above case, x was a single-state variable, that is, a scalar quantity. However, when many state variables are considered together, then $\mathbf{x}$ is a vector, $\mathbf{A}$ is a matrix and the parabola becomes a paraboloid. However, the minimum still occurs where $\mathbf{Ax} = \mathbf{b}$ if $\mathbf{A}$ is positive definite.
In the n-dimensional case, the energy equation is written as

$$E_{\mathrm{N}} = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{b} \tag{11.12}$$

and if $\mathbf{A}$ is positive definite, the minimum occurs at

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{11.13}$$

11.4 Least Squares Solution

Now reverting back to the earlier problem of error minimization where the error quadratic is as given in Eq. (11.5) and is repeated here for clarity.

$$f(x) = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{e}^{\mathrm{T}}\mathbf{e}$$

The error in terms of the state vector is given as

$$\begin{aligned} e_1 &= y_{\mathrm{act}1} - y_1 = y_{\mathrm{act}1} - \mathbf{C}\mathbf{x} \\ e_2 &= y_{\mathrm{act}2} - y_2 = y_{\mathrm{act}2} - \mathbf{C}\mathbf{x} \\ &\vdots \\ e_n &= y_{\mathrm{act}n} - y_n = y_{\mathrm{act}n} - \mathbf{C}\mathbf{x} \end{aligned}$$

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} y_{\text{act}1} \\ y_{\text{act}2} \\ \vdots \\ y_{\text{act}n} \end{bmatrix} - \begin{bmatrix} \mathbf{C} \\ \mathbf{C} \\ \vdots \\ \mathbf{C} \end{bmatrix} \mathbf{x}$$

This can be written as

$$\mathbf{E} = \mathbf{Y}_{\mathbf{act}} - \mathbb{C}\mathbf{x} \tag{11.14}$$

where $\mathbf{Y}_{\mathbf{act}}$ is the vector of the actual or measured values of the output and $\mathbf{x}$ is the state vector. The quadratic error function f can be written as

$$\begin{aligned} f &= \mathbf{E}^{\mathrm{T}}\mathbf{E} = (\mathbf{Y}_{\text{act}} - \mathbb{C}\mathbf{x})^{\mathrm{T}}(\mathbf{Y}_{\text{act}} - \mathbb{C}\mathbf{x}) \\ &= (\mathbb{C}\mathbf{x} - \mathbf{Y}_{\text{act}})^{\mathrm{T}}(\mathbb{C}\mathbf{x} - \mathbf{Y}_{\text{act}}) \\ &= \mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\,\mathbb{C}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} - \mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\,\mathbb{C}\mathbf{x} + \mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} \end{aligned}$$

From the above, it is evident that

$$\mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} = (\mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\,\mathbb{C}\mathbf{x})^{\mathrm{T}}$$

Therefore, the function f can be equivalently written as

$$f = \mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\,\mathbb{C}\mathbf{x} - 2\mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\,\mathbb{C}\mathbf{x} + \mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} = \mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\,\mathbb{C}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} - \mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}}$$

From the above, it is evident that f has two components, $J_1(\mathbf{x})$ which is a function of $\mathbf{x}$ and J_2 which is not a function of $\mathbf{x}$. Thus,

$$f(x) = J_1(\mathbf{x}) - J_2$$

where

$$J_1(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\,\mathbb{C}\mathbf{x} - 2\mathbf{x}^{\mathrm{T}}\,\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} \text{ and } J_2 = \mathbf{Y}_{\mathbf{act}}{}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}}$$

Evidently, minimizing f with respect to x means that one minimizes $J_1(\mathbf{x})$ with respect to $\mathbf{x}$ as J_2 is not a function of $\mathbf{x}$. Therefore, it is sufficient to only consider $J_1(\mathbf{x})$ for purposes of minimization. Thus,

$$J = \frac{J_1}{2} = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbb{C}^{\mathrm{T}}\mathbb{C}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} \tag{11.15}$$

Therefore, minimizing J implies that f is minimized. Now compare Eq. (11.15) with the energy equation [Eq. (11.12)]. J and E_{N} are of similar form wherein

$$\mathbf{x} \rightarrow \mathbf{x}$$

$$\mathbb{C}^{\mathrm{T}}\,\mathbb{C} \rightarrow \mathbf{A}$$

$$\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} \rightarrow \mathbf{b}$$

A minimum for J would exist if $\mathbb{C}^{\mathrm{T}}\,\mathbb{C}$ is positive definite. Equation (11.13) gives the solution of $\mathbf{x}$ such that E_{N} is minimum. Applying the analogous equivalence as shown above to the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ that gives the minimum, one obtains

$$\mathbf{x} = (\mathbb{C}^{\mathrm{T}}\mathbb{C})^{-1}\mathbb{C}^{\mathrm{T}}\mathbf{Y}_{\mathbf{act}} \tag{11.16}$$

Thus, Eq. (11.16) is called the least squares solution of $\mathbf{E} = \mathbf{Y}_{\mathbf{act}} - \mathbb{C}\mathbf{x}$. This means that the value of the state vector $\mathbf{x}$ as decided by Eq. (11.16) will give the smallest or the minimum square error when used in the output equation model $\mathbf{y} = \mathbf{Cx}$.

Problem 11.1

Consider the following deterministic system:

$$(1 + b_1 z^{-1}) y_k = (a_0 + a_1 z^{-1}) u_{k-1}$$

Formulate an estimator to estimate a_0, a_1 and b_1 using the least square algorithm for minimizing the error.

Solution

Problem formulation

$$y_k = a_0 u_{k-1} + a_1 u_{k-2} - b_1 y_{k-1}$$

$$y_k = [u_{k-1} \quad u_{k-2} \quad -y_{k-1}] \begin{bmatrix} a_0 \\ a_1 \\ b_1 \end{bmatrix}$$

$$y = \mathbf{Cx}$$

where

$$\mathbf{C} = [u_{k-1} \quad u_{k-2} \quad -y_{k-1}]$$

$$\mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \\ b_1 \end{bmatrix}$$

Problem 11.2

Consider the system

$$(1 + b_1 z^{-1} + b_2 z^{-2}) y_k = (a_0 + a_1 z^{-1} + a_2 z^{-2}) u_{k-1}$$

Formulate the above system as a least square estimation problem to estimate the parameters a_0, a_1, a_2, b_1 and b_2.

Solution

$$y_k = a_0 u_{k-1} + a_1 u_{k-2} + a_2 u_{k-3} - b_1 y_{k-1} - b_2 y_{k-2}$$

$$y_k = [u_{k-1} \quad u_{k-2} \quad u_{k-3} \quad -y_{k-1} \quad -y_{k-2}] \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}$$

$$y = \mathbf{Cx}$$

where

$$\mathbf{C} = [u_{k-1} \quad u_{k-2} \quad u_{k-3} \quad -y_{k-1} \quad -y_{k-2}]$$

$$\mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}$$

11.5 Weighted Least Squares

It is seen in the previous section that for a set of measured values represented by the vector $\mathbf{y}_{\mathbf{act}}$, the best estimate of the state vector $\mathbf{x}$ is given by the least square solution as given in Eq. (11.16). If now fresh values of the output are measured, then $\mathbf{y}_{\mathbf{act}}$ vector will get updated. The best estimate of $\mathbf{x}$ should again be recalculated. It can be seen from Eq. (11.16) that the accuracy of the best estimate of $\mathbf{x}$ is solely dependent on the accuracy with which the output is measured or observed. This means that $\mathbf{x}$ is as good as one can trust the values of $\mathbf{y}_{\mathbf{act}}$. As measurements continue to arrive, the estimate of $\mathbf{x}$ is obtained from the data by a straight line fit. For each new data point there will be a change in the best line. To obtain the best estimate of $\mathbf{x}$ in the face of measurement and observation errors, one must include a *trust factor or weighting factor* for the error due to measurements. The trust factor or the weighting factor for a batch of measurements can be decided qualitatively on the basis of various factors such as sensing methods, temperature changes during sensing, operator fatigue, etc. Thus, if $\mathbf{e}$ is the error vector, then from Eq. (11.4), one has

$$\mathbf{e} = (\mathbf{y}_{\mathbf{act}} - \mathbb{C}\mathbf{x}) = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{11.17}$$

Let $w_1, w_2, \ldots, w_n$ be the weighting factors or trust factors for the errors $e_1, e_2, \ldots, e_n$ respectively. Then the weighted errors are represented as

$$\begin{bmatrix} w_1 e_1 \\ w_2 e_2 \\ \vdots \\ w_n e_n \end{bmatrix} = \begin{bmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & w_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbb{W}\mathbf{e} \tag{11.18}$$

where $\mathbb{W}$ is a diagonal matrix containing the trust factors as its diagonal elements and $\mathbf{e}$ is the error vector. Pre-multiplying Eq. (11.17) by the weight matrix $\mathbb{W}$, one obtains

$$\mathbb{W}\mathbf{e} = (\mathbb{W}\mathbf{y}_{\mathbf{act}} - \mathbb{W}\mathbb{C}\mathbf{x}) \tag{11.19}$$

It is now needed to find a solution to the above equation such that the weighted square error is minimized. This means that $(\mathbf{We})^{\mathrm{T}}(\mathbf{We})$ should be minimized. The least square solution of $\mathbf{e} = \mathbf{y}_{\mathbf{act}} - \mathbb{C}\mathbf{x}$ is as given

in Eq. (11.16). Comparing Eq. (11.19) with $\mathbf{e} = \mathbf{y}_{\mathbf{act}} - \mathbb{C}\mathbf{x}$, the least square solution can be obtained by the following analogous equivalence:

$$\mathbf{e} \rightarrow \mathbb{W}\mathbf{e}$$
$$\mathbf{y}_{\mathbf{act}} \rightarrow \mathbb{W}\mathbf{y}_{\mathbf{act}}$$
$$\mathbb{C} \rightarrow \mathbb{W}\,\mathbb{C}$$
$$\mathbf{e}^{\mathbf{T}} \rightarrow \mathbf{e}^{\mathbf{T}}\mathbb{W}^{\mathbf{T}}$$
$$\mathbb{C}^{\mathbf{T}} \rightarrow \mathbb{C}^{\mathbf{T}}\mathbb{W}^{\mathbf{T}}$$

Using the above equivalence and applying them in Eq. (11.16), one obtains

$$\mathbf{x} = (\mathbb{C}^{\mathrm{T}}\mathbb{W}^{\mathrm{T}}\mathbb{W}\mathbb{C})^{-1}\mathbb{C}^{\mathrm{T}}\mathbb{W}^{\mathrm{T}}\mathbb{W}\mathbf{y}_{\mathbf{act}} \tag{11.20}$$

Equation (11.20) is the weighted least square solution for Eq. (11.19). The term $\mathbb{W}^{\mathrm{T}}\mathbb{W}$ that occurs in the weighted least square solution of Eq. (11.19) can be written as

$$\mathbb{W}^{\mathrm{T}}\mathbb{W} = \begin{bmatrix} {w_1}^2 & 0 & 0 & 0 \\ 0 & {w_2}^2 & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & {w_n}^2 \end{bmatrix} \tag{11.21}$$

As the measured data arrive on each measurement time interval, the variance also varies according to the trust. If the measured data are wild, then the variance is large and vice-versa. If w_i is set equal to $1/\sigma_i^2$ where σ_i^2 is the variance at the ith measurement interval, then the weighting factor gives a measure of the quality of the observed values. If the variance is large then w_i is small and hence that particular observed value is given less importance. Therefore, variance can be used for deciding the weighting factors or trust factors.

11.6 Recursive Least Squares

As the measurements arrive and the straight line fit is continuously updated. For each new measurement or observation of $\mathbf{y}_{\mathbf{act}}$, the new best line has to be found out. This means that the best value of vector $\mathbf{x}$ has to be estimated for every new update in the measurement such that the latest solution of $\mathbf{x}$ is the least square solution. This implies that for every new update, the least square solution for $\mathbf{x}$ has to be recalculated. There are two major drawbacks in this approach: (a) the $\mathbf{y}_{\mathbf{act}}$ vector size continuously increases as more and more data arrive and (b) the least square solution for $\mathbf{x}$ needs to be calculated for every measurement update which will become more and more time consuming as $\mathbf{y}_{\mathbf{act}}$ vector length grows.

Therefore to overcome the above drawbacks, recursive techniques are used. In this approach, the newest best estimate of vector $\mathbf{x}$ is dependent on the previous best estimate and the latest measurement. This can be represented as

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \text{Correction} \tag{11.22}$$

where $\mathbf{x}(k + 1)$ is the estimate of state vector at $(k + 1)$th iteration; $\mathbf{x}(k)$ is the estimate of the state vector at the kth iteration and the correction is dependent on the newest error due to the latest measurement $\mathbf{y}_{\mathbf{act}}$. At this point it is now appropriate to introduce the following changes in notation to distinguish between certain variables. Therefore,

1. $\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}_k$ is the estimate of the state vector at kth iteration. The actual state vector at kth iteration in the physical system is denoted by $\mathbf{x}(k)$ or $\mathbf{x}_k$.
2. $\mathbf{y}(k) = \mathbf{y}_k = \mathbf{y}_{\text{act}}$ is the measurement or observed value of the output at the kth iteration.

Now the general recursive equation [Eq. (11.22)] can be re-written as

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \text{Correction}_k \tag{11.23}$$

The value of Correction_k is proportional to the error due to the latest measurement which will take corrective action on the value of the previous best estimate of the state vector to give the latest best estimate of the state vector. Thus, the recursive estimate equation becomes

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + Ge_k = \hat{\mathbf{x}}_k + G(y_k - \mathbf{C}\hat{\mathbf{x}}_k) \tag{11.24}$$

where e_k is the error at the kth iteration. Equation (11.24) is the general recursive estimator equation. G is called the correction gain. In fact, the various algorithms differ in the way in which G is chosen. There are many algorithms that use the above recursive structure; however, this section will discuss two of the more popular algorithms: the least mean square (LMS) algorithm and the Kalman filter algorithm. The LMS algorithm chooses G in such a way that the minimum mean square error is reached by descending to the bottom of the parabola with the steepest possible slope. The Kalman filter algorithm chooses G in such a way that the error in the presence of system noise is minimized. In all the cases, the result is the best estimate of the state vector that will minimize some cost function J.

LMS Algorithm

The least mean square algorithm is also called LMS algorithm. This is also very popular in adaptive filtering applications such as noise cancellation apart from its application in estimators. In the previous sections, the optimization was done with respect to the square of the sum of errors or the quadratic of the error vector, that is, $\mathbf{e}^{\mathrm{T}}\mathbf{e}$. Here, as the name suggests, the optimization is done with respect to the mean of the square of sum of errors or mean/expectation of the quadratic of the error vector, that is, $E(\mathbf{e}^{\mathrm{T}}\mathbf{e})$. Let J be the function which is equal to the mean/expectation of the error quadratic. Then one has to find a value of the state vector $\hat{\mathbf{x}}$ such that J is a minimum. Such an $\mathbf{x}$ will be the optimal solution or the optimal estimate of the state vector. The function to minimize J is shown in Figure 11.4 as a parabola for the case of a one-dimensional state vector x. Here the state variable value $x_{\min}$ corresponds to the minimum value of J, that is, $J_{\min}$.

The sequence of finding the state variable value for the one-dimensional case as shown in Figure 11.4 is illustrated in Figure 11.5. Consider that at an arbitrary iteration k, the state variable is x_k and the corresponding cost function value is J_k. The J_k point is shown on the parabola in Figure 11.5(a). At the point J_k on the parabola, grad J_k or ∇J_k is the slope of J_k with respect to x. This slope is used as the correction term for the next update of the state variable. To approach the minimum of J, one would like to take the opposite direction to descend quickly or in other words the steepest descent. This is done by subtracting the present value of the state variable with the present scaled slope of J to obtain the next value of the state variable. Thus the recursive estimate equation is

$$\hat{x}_{k+1} = \hat{x}_k + \mu(-\nabla J_k) \tag{11.25}$$

Equation (11.25) is the algorithm that will iteratively find the best estimate of the state vector such that the cost function J is a minimum. The negative sign for ∇J_k in Eq. (11.25) indicates that the opposite direction to the slope at the point is taken to provide fastest descent to the minimum.

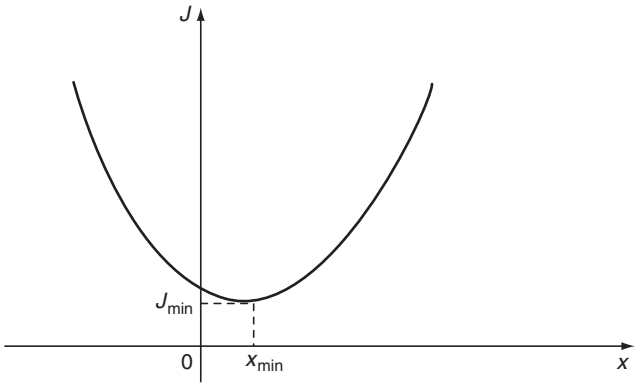


Figure 11.4 | One-dimensional state vector versus J.

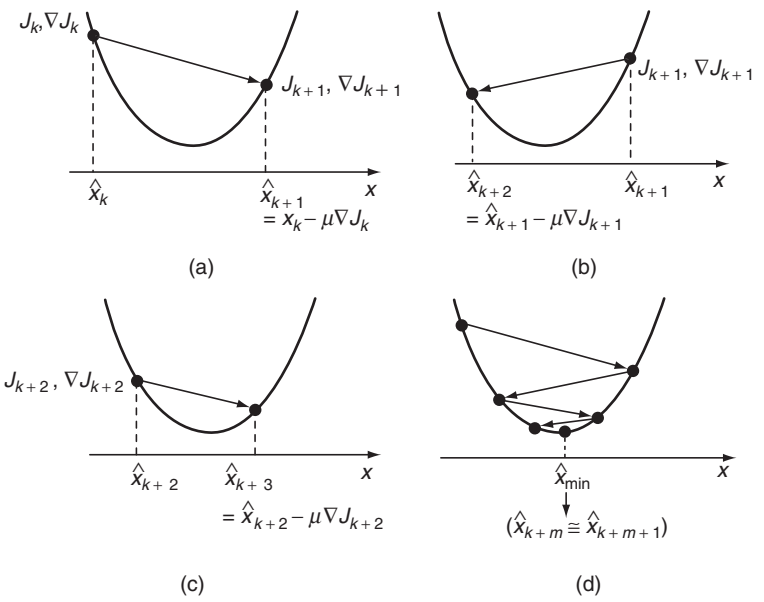


Figure 11.5 | Steepest descent technique for the LMS algorithm.

The variable μ in Eq. (11.25) is a positive number called the slope scaling factor. This is also called the learning rate. For now, let μ be appropriately chosen such that J converges. Referring to Figure 11.5(a), the slope at J_k is negative. For minimizing one should move in the opposite direction. Therefore, a scaled positive slope value is added to the previous value of the state estimate. The next state variable estimate is always to the right of the previous state variable estimate if the previous slope of J is negative. $\hat{x}_{k+1}$ is the new state variable estimate. The cost function value now is J_{k+1} and the slope at this point is ∇J_{k+1} which is now

positive. The negative of this positive slope is scaled by μ and added to the present state variable estimate to obtain the next state variable estimate $\hat{x}_{k+2}$. The next state variable estimate is always to the left of the present state variable estimate if the present slope of J is positive. In this way, after a few iterations as illustrated in Figure 11.5(d), the cost function will converge to its minimum value $J_{\min}$ at which point the state variable estimate is $x_{\min}$ which is the optimal solution. At the minimum, the previous and present state variable estimates will be almost equal when the iteration can be stopped.

It should be noted that the choice of the slope scaling factor or the learning rate μ is important to ensure stability of the minimization of the process (Figure 11.6). If the value of μ is higher than a critical value μ_{cri}, then J_{k+1} will be higher than J_k and therefore will not converge. This is an unstable situation wherein the minimum cannot be found. If μ is lower than the value of μ_{cri} then J_{k+1} will be lower than J_k and will eventually converge to the minimum. This is the stable situation. One should also note that if μ is very much lower than μ_{cri}, then the convergence is very slow, that is, the learning rate is slow. Therefore, many versions of the LMS algorithm exist strictly by the choice of the learning rate. However, as a simple practical approach, one may start the minimization process with a very low value of μ and then tune μ by increasing it to improve the learning rate.

Equation (11.25) is not an easy equation to implement as ∇J_k is difficult to calculate in every iteration. Therefore, it is necessary to obtain an approximation for ∇J_k that is an easily implementable measure of the slope of J at iteration k. For the one-dimensional case, it is known that

$$J_k = \left(y_k - C\hat{x}_k\right)^2$$

$$\nabla J_k = \frac{\partial J_k}{\partial \hat{x}_k} \approx (-C)2(y_k - C\hat{x}_k)$$

where the above relationship is shown as an approximation. This is because even though y_k is a measured value, it is dependent on $\hat{x}_k$ through the controller which will affect the measurements. However, in the above relationship this dependency is ignored. Thus,

$$\nabla J_k \approx -2Ce_k \tag{11.26}$$

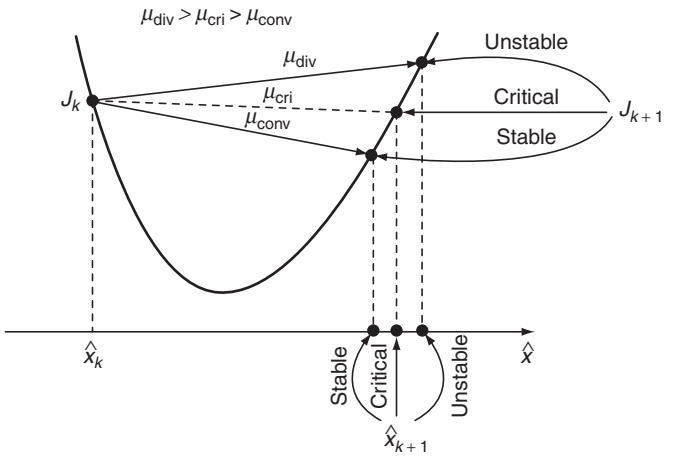


Figure 11.6 | Choice of slope scaling factor or learning rate μ.

Substituting Eq. (11.26) in Eq. (11.25), one obtains

$$\hat{x}_{k+1} = \hat{x}_k + 2\mu C e_k \tag{11.27}$$

Equation (11.27) is the implementable LMS recursive algorithm for the one-dimensional case. Comparing with Eq. (11.24), it is evident that the estimator gain G is equal to $2\mu C$ for the LMS recursive algorithm. The LMS algorithm for the one-dimensional case can be extended to the multi-dimensional case along-similar lines. Thus for the multi-dimensional case,

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + 2\mu \mathbf{C}^{\mathrm{T}} e_k \tag{11.28}$$

where e_k is the error at the kth iteration.

Problem 11.3

Consider the following deterministic system,

$$(1 + b_1 z^{-1}) y_k = (a_0 + a_1 z^{-1}) u_{k-1}$$

Formulate an estimator to estimate a_0, a_1 and b_1 using the LMS algorithm for minimizing the mean square error.

Solution

Formulation of the LMS problem:

$$y_k = a_0 u_{k-1} + a_1 u_{k-2} - b_1 y_{k-1}$$

$$y_k = \begin{bmatrix} u_{k-1} & u_{k-2} & -y_{k-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ b_1 \end{bmatrix}$$

$$y_k = \mathbf{C}\mathbf{x}$$

where

$$\mathbf{C} = \begin{bmatrix} u_{k-1} & u_{k-2} & -y_{k-1} \end{bmatrix}$$

and

$$\mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \\ b_1 \end{bmatrix}$$

Here $\mathbf{x}$ has to be estimated. The estimate equation is given as

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + 2\mu \mathbf{C}^{\mathrm{T}} e_k$$

where μ is the learning rate and $e_k = y_{\mathrm{act}} - \mathbf{C}\hat{\mathbf{x}}_k$.
The above formulation is simulated in MATLAB and the tracking of the estimate vector $\mathbf{x}$ for different learning rates are plotted in Figures 11.7–11.9.
It can be observed that high learning rate requires lesser number of iterations to estimate the states.

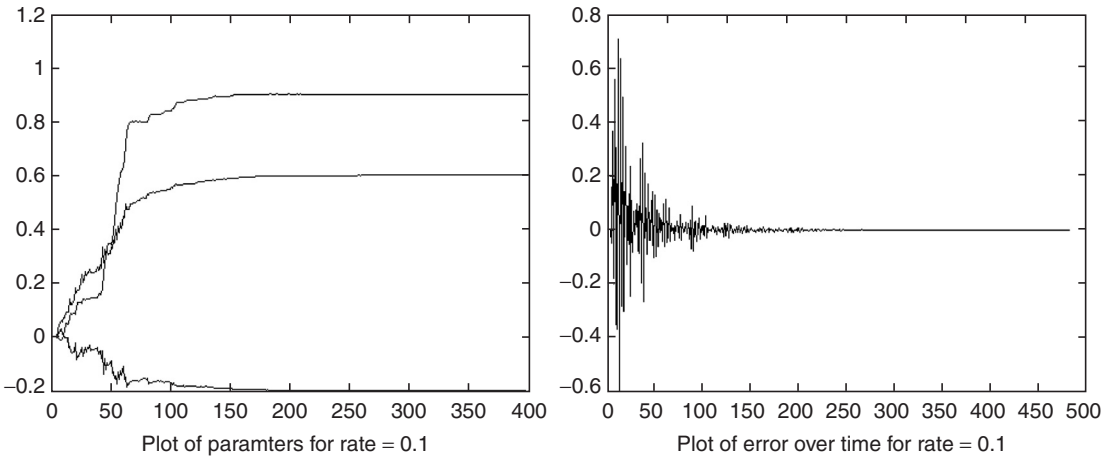


Figure 11.7 | Plot of **x** and *e* for learning rate $\mu = 0.1$ – $\mathbf{x}^T = [0.6\ -0.2\ 0.9]$.

Kalman Filter

Kalman filter is an optimal estimator. It is also a recursive algorithm that is used for the optimal estimation of the states. This algorithm gives the optimal state estimates taking into account the noise characteristics of the system also. Till now, the state space model of the physical systems did not consider the noise of the system in the state equation and the noise of the measuring or sensing devices in the output equation. If noise is also considered then the system state space model is written as

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k \tag{11.29}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \tag{11.30}$$

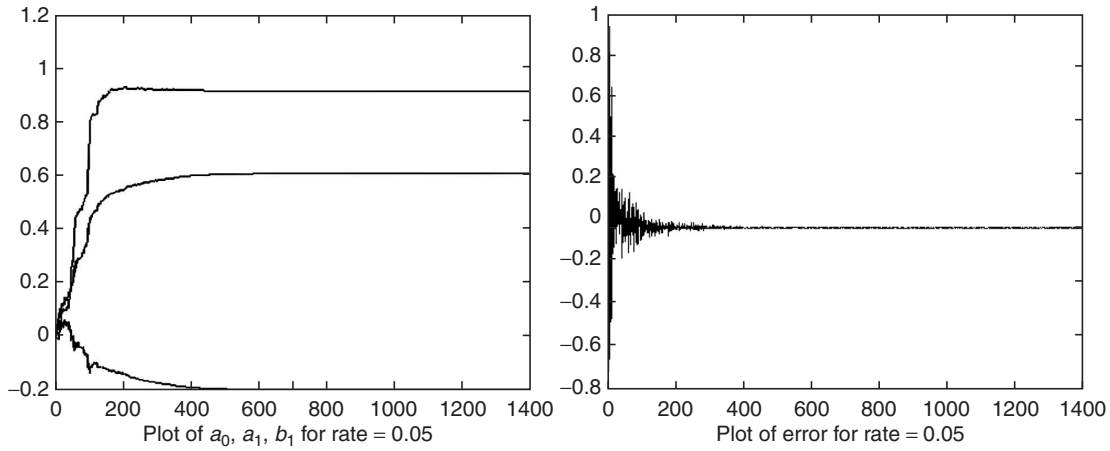


Figure 11.8 | Plot of **x** and *e* for learning rate $\mu = 0.05$ – $\mathbf{x}^T = [0.6\ -0.2\ 0.9]$.

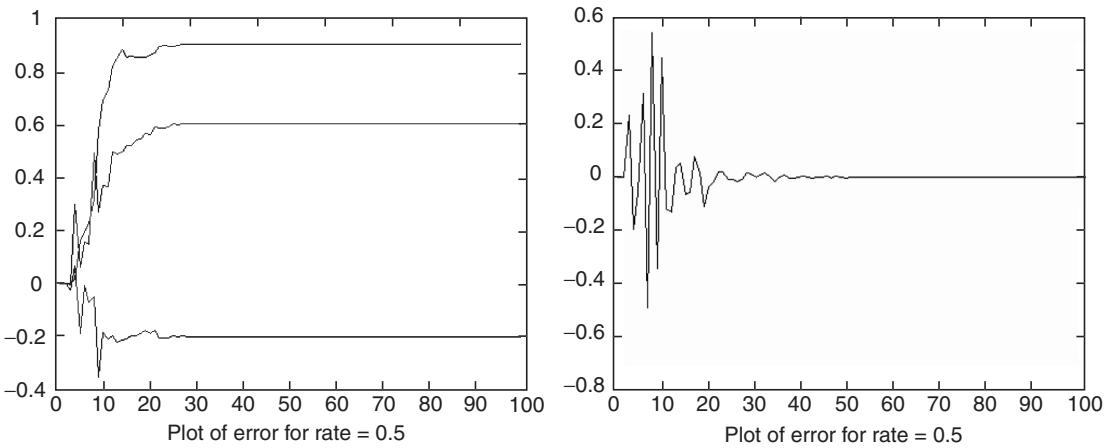


Figure 11.9 | Plot of **x** and *e* for learning rate $\mu = 0.5 - \mathbf{x}^{\mathsf{T}} = [0.6\ -0.2\ 0.9]$.

where $\mathbf{w}_k$ is the plant noise and $\mathbf{v}_k$ is the sensor or the measurement noise. In LMS algorithm, it is seen that the next state update is given by a one-step *time update* equation [Eq. (11.28)]. However, in the case of the Kalman filter, there are two steps involved. First the *time update* is done and then a *measurement update* is performed on the time updated state variable. The recursion steps for the Kalman filter are as follows:

Step 1: Time update

$$\tilde{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k \tag{11.31}$$

where $\tilde{\mathbf{x}}_{k+1}$ is an intermediate state variable estimate which is obtained from the system state equation.

Step 2: Measurement update

$$\hat{\mathbf{x}}_{k+1} = \tilde{\mathbf{x}}_{k+1} + \mathbf{G}_{k+1}(\mathbf{y}_{k+1} - \mathbf{C}\tilde{\mathbf{x}}_{k+1}) \tag{11.32}$$

where $\mathbf{G}_{k+1}$ is the estimator gain or the gain of the Kalman filter.

Observe that the Kalman filter is based on the current estimator concept of a time update followed by a measurement update. During the time update, the estimate of the kth iteration is used to determine an intermediate state vector of the $(k + 1)$th iteration, $\tilde{\mathbf{x}}_{k+1}$. During the measurement update, the update action is taken on variables that are all in the same time iteration level. The latest measurement $\mathbf{y}_{k+1}$ is used to take correction action to obtain the final estimate $\hat{\mathbf{x}}_{k+1}$ at the $(k + 1)$th iteration.

To implement the Kalman filter algorithm, there are a few stochastic terms that need to be described due to the presence of noise terms in the model. The noises need to be characterized by two measures: (a) the mean or the expected value and (b) the co-variance.

The noises $\mathbf{w}_k$ and $\mathbf{v}_k$ have a mean or average value that is zero. This means that the expected values of $\mathbf{w}_k$ and $\mathbf{v}_k$ are zero. Thus,

$$E(\mathbf{w}_k) = E(\mathbf{v}_k) = 0$$

The co-variance is the expected or the mean of the quadratic. The co-variance gives a measure of the energy content in the noise. Thus, $E(\mathbf{w}_k^{\mathrm{T}}\mathbf{w}_k)$ is called the co-variance of noise w_k. Likewise, $E(\mathbf{v}_k^T\mathbf{v}_k)$ is the co-variance of noise $\mathbf{v}_k$. In general, any signal is characterized as

$$\text{signal} \rightarrow \{\text{mean, co-variance}\}$$

Both the noise measures will have to be found out by repeated measurements on the physical system. For the system or plant noise $\mathbf{w}_k$, the mean is zero and the co-variance is some value $\mathbf{Q}$ that is greater than or equal to zero. Then,

$$\mathbf{w}_k \rightarrow \{0, \mathbf{Q}\} \text{ qualifies the plant noise}$$

For the measurement noise $\mathbf{v}_k$, the mean is again zero and the co-variance is some value $\mathbf{R}$ that is greater than zero. Then, $\mathbf{v}_k$ is characterized as

$$\mathbf{v}_k \rightarrow \{0, \mathbf{R}\} \text{ which qualifies the sensor or measurement noise, } \mathbf{v}_k$$

Similarly, the state vector can also be characterized in terms of the mean and the co-variances. However, the mean value for the state vector is not zero. Let the mean value of the state vector be designated as $\overline{\mathbf{x}}_k$ which is the mean value in a given interval kT and the co-variances are taken about the mean value. Thus, one can define

$$\mathbf{P}_k = E\{(\mathbf{x}_k - \overline{\mathbf{x}}_k)^{\mathrm{T}}(\mathbf{x}_k - \overline{\mathbf{x}}_k)\}$$

which is the co-variance of the state variable value taken about the mean or in other words, it can be called the error co-variance as it indicates the co-variance of the deviation of the state about the mean value in that interval. Thus, the state vector can be characterized as

$$\mathbf{x}_k \rightarrow \{\overline{\mathbf{x}}_k, \mathbf{P}_k\}$$

It should be noted that the cross co-variances between $\mathbf{w}_k$, $\mathbf{v}_k$ and the state vector are zero. This means that $\mathbf{w}_k$, $\mathbf{v}_k$ and the state vector $\mathbf{x}$ occur independent of each other. Thus,

$$E(\mathbf{w}^{\mathrm{T}}\mathbf{x}) = E(\mathbf{w}^{\mathrm{T}}\mathbf{x}) = E(\mathbf{v}^{\mathrm{T}}\mathbf{x}) = 0$$

With this brief background discussed above in characterizing signals based on means and co-variances, the Kalman filter algorithm can be understood and implemented. The state vector error co-variance $\mathbf{P}_k$ is the important parameter of concern in this algorithm. $\mathbf{P}_k$ is the quadratic of the deviation of the state vector value about its expected value. Thus $\mathbf{P}_k$ is the state error quadratic or the state error co-variance. Here $\mathbf{P}_k$ needs to be minimized. Therefore, the propagation of $\mathbf{P}_k$ across each iteration step is critical. $\mathbf{P}_k$ should ultimately settle down to a minimum. At the (zero)th iteration, the state vector is characterized as $\mathbf{x}_0 \rightarrow \{\overline{\mathbf{x}}_0, \mathbf{P}_0\}$ with some initial value of mean of state and error co-variance. Figure 11.10 illustrates the propagation of the various variables across two arbitrary iteration steps k and $k+1$.

Figure 11.10 illustrates the process of Kalman filtering. The ultimate goal of Kalman filter is to obtain the next estimate of the state vector such that the error co-variance converges to a minimum. The associated error correction gain $\mathbf{G}_{k+1}$ is called the Kalman filter gain or the Kalman estimator gain. Three iteration steps are shown in Figure 11.10. At the end of the kth iteration, the state error co-variance $\mathbf{P}_k$ and the state estimate $\hat{\mathbf{x}}_k$ are available. First the state error co-variance is updated in time called the time updated value which is an intermediate variable $\tilde{\mathbf{P}}_{k+1}$. Without going into the derivations, as these derivations are out of the scope of the book, the time updated state error co-variance is given as

$$\tilde{\mathbf{P}}_{k+1} = \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{Q} \tag{11.33}$$

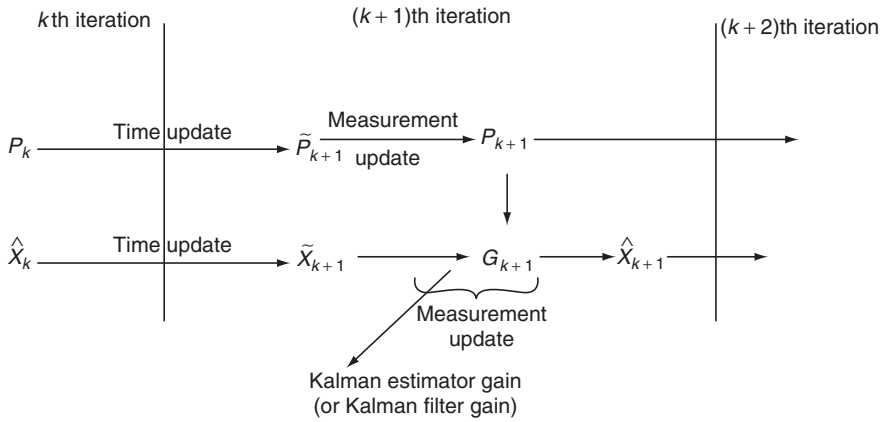


Figure 11.10 | Time update and measurement update in Kalman filter algorithm.

where **A** is the system matrix as given in the system state equation and **Q** is the system or plant noise co-variance as described earlier.

The measurement noise will also affect the state error. The measurement noise co-variance **R** will be used to update the time updated state error co-variance. This is called the measurement update for the state error co-variance which will result in the state error co-variance value in the $(k+1)$th iteration. Thus the measurement update of the state error co-variance is given by

$$\mathbf{P}_{k+1} = \tilde{\mathbf{P}}_{k+1} - \tilde{\mathbf{P}}_{k+1}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\tilde{\mathbf{P}}_{k+1}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{C}\tilde{\mathbf{P}}_{k+1} \tag{11.34}$$

Note that the measurement update is dependent on only $\tilde{\mathbf{P}}_{k+1}$, $\mathbf{C}$ and $\mathbf{R}$. The state error co-variance of the $(k+1)$th iteration as given in Eq. (11.34) is used for calculating the $(k+1)$th iteration Kalman filter gain $\mathbf{G}_{k+1}$, which is given as

$$\mathbf{G}_{k+1} = \mathbf{P}_{k+1}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1} \tag{11.35a}$$

The time update for the state vector using the state vector estimate of the kth iteration is given in Eq. (11.31) which is repeated here for clarity:

$$\tilde{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k$$

where $\tilde{\mathbf{x}}_{k+1}$ is a intermediate time updated state vector estimate which is obtained from the system state equation. Now using the latest measurements obtained in the $(k+1)$th iteration, the latest correction can be applied to the time updated state vector estimate to obtain the state vector estimate of the $(k+1)$th iteration. The measurement update of the state vector estimate is given by Eq. (11.32) which is repeated here for clarity:

$$\hat{\mathbf{x}}_{k+1} = \tilde{\mathbf{x}}_{k+1} + \mathbf{G}_{k+1}(\mathbf{y}_{k+1} - \mathbf{C}\tilde{\mathbf{x}}_{k+1})$$

It should be noted that Eq. (11.35a) is not the only equation for Kalman filter gain. There are other variants also.

Steady-State Kalman Fitler

Under steady-state conditions, $\mathbf{P}_{k+1} = \tilde{\mathbf{P}}_{k+1} = \mathbf{P}_k = \tilde{\mathbf{P}}_k = \mathbf{P}$. Therefore, the Kalman gain from Eq. (11.35a) is given as

$$\mathbf{G} = \mathbf{P}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1} \tag{11.35b}$$

The gain obtained from Eq. (11.35b) is called the steady-state Kalman gain. This will, however, give sub-optimal estimator performance.

The implementation algorithm for the Kalman filter is discussed next.

Choose $\mathbf{P}_o = \mathrm{I}$ and $\mathbf{x}_o = \mathbf{0}$ and evaluate the estimator equations in the following sequence:

1. $\mathbf{P}_{k+1} = \mathbf{A}\mathbf{P}_k\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$
2. $\mathbf{P}_{k+1} = \tilde{\mathbf{P}}_{k+1} - \tilde{\mathbf{P}}_{k+1}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\tilde{\mathbf{P}}_{k+1}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{C}\tilde{\mathbf{P}}_{k+1}$
3. $\mathbf{G}_{k+1} = \mathbf{P}_{k+1}\mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}$
4. $\tilde{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k$
5. $\hat{\mathbf{x}}_{k+1} = \mathbf{A}\tilde{\mathbf{x}}_{k+1} + \mathbf{G}_{k+1}(\mathbf{y}_{k+1} - \mathbf{C}\tilde{\mathbf{x}}_{k+1})$

11.7 Optimal Control: Linear Quadratic

Referring to the system block diagram as shown in Figure 11.11, the optimal estimation has been the focus of discussion until now. The principle of minimum energy is used to obtain the optimal estimator gain **G** such that a cost function based on the square of the error quadratic is minimized. This very same principle of minimum energy can be used to obtain the optimal controller gain **K** such that a specific cost function is minimized. The previous chapter discusses the concepts and principles of the regulator problem and the tracker problem including the design of controllers and estimators for such structures. Here the algorithm for obtaining the optimal controller gain **K** will be addressed.

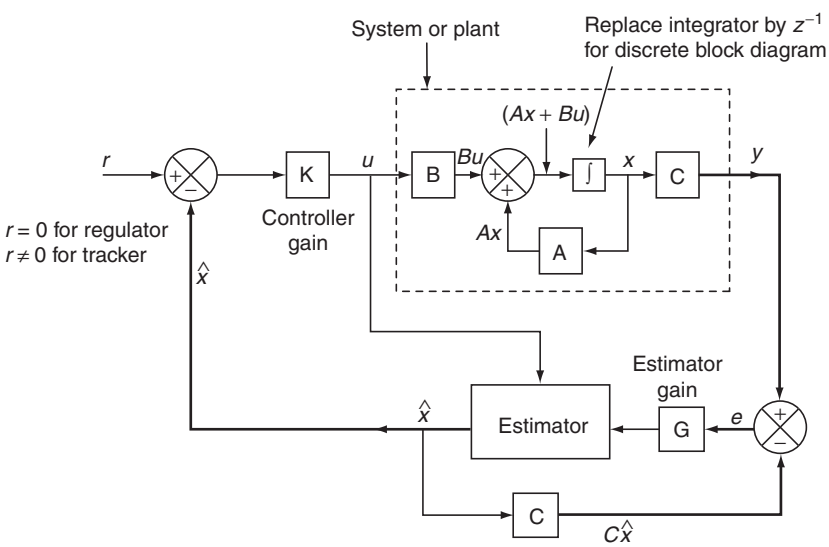


Figure 11.11 | Controller and estimator for a given system using state space control methods.

In the case of the regulator problem where the reference input is zero, all the states will ultimately settle to zero in full-state feedback configuration. This means that the input which is proportional to the state values will also settle down to zero in the steady state. On the occurrence of a disturbance one would like the states to quickly settle down back to zero. One could minimize the area under the square of the state deviation from zero and/or the area under the square of the input deviation from zero. This is illustrated in Figure 11.12. Figure 11.12 shows the one-dimensional case for x and u. The time evolution of x and u on the occurrence of a disturbance is shown. Ultimately x and u will settle to zero. The time evolutions of x^2 and u^2 are also shown. The area under the x^2 and u^2 curves should be minimized so that the response will match the reference as quickly as possible. A possible candidate cost function J can be taken as the mean of the area under the x^2 and u^2 curves. Thus,

$$J = \frac{\text{area}_u + \text{area}_x}{2}$$

Referring again to Figure 11.12, and substituting for the area_u and area_x, one obtains

$$J = \frac{1}{2}\int_0^\infty (u^2 + x^2)\mathrm{d}t$$

The above cost function does not give much design flexibility as the input quadratic and the state quadratic are not weighted. One may weigh the input quadratic by an appropriate weight R and the state quadratic by another appropriate weight Q. Then the modified cost function is given as

$$J = \frac{1}{2}\int_0^\infty (Ru^2 + Qx^2)\mathrm{d}t$$

Q and R can be used as the design parameters in the minimization process to tune the value of the controller gain K. By adjusting the values of Q and R, one can obtain the controller gain values such that the response is faster or slower.

For the multi-dimensional case where x and u are vectors, then Q and R become weighting matrices, respectively. Thus, for the general multi-dimensional system, the cost function is given as

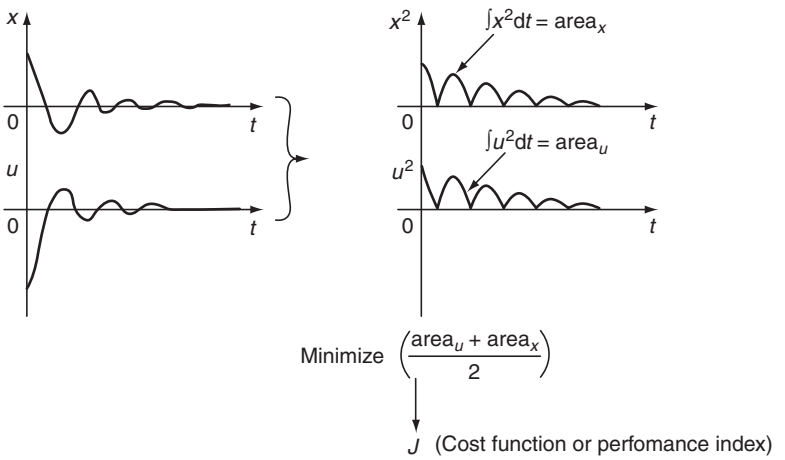


Figure 11.12 | Cost function for optimal control.

$$J = \frac{1}{2}\int_0^\infty (\mathbf{u}^T\mathbf{R}\mathbf{u} + \mathbf{x}^T\mathbf{Q}\mathbf{x})\mathrm{d}t \tag{11.36}$$

where **Q** is a diagonal matrix containing the weighting factors for each state and **R** is a diagonal matrix containing the weighting factors for each input. (Note that the Q and R used here are different from the noise co-variance matrices Q and R used in the Kalman estimator.) J is a quadratic of both the input vector and the state vector.

The cost function J in Eq. (11.36) consists of two quadratic forms. One is the quadratic of the state vector and the other is the quadratic of the input vector. The objective is to design or select the controller gains **K** such that the minimum sufficient input **u** is applied to obtain the specific performance objective. This means that the controller gains should be selected such that the input quadratic or the input energy is minimized. One can add one more constraint by including the minimization of the state energy. This would imply that the controller gains be selected so that the input energy as well as the state energy is minimized. Then such a choice of gain **K** is called the optimal controller. The states are in general controllable from the input. In such a case, minimizing only the input energy is sufficient. Therefore, **Q** can be either zero or positive, that is, **Q** should be positive semi-definite. However, **R** should be positive definite for a minimum to exist.

It should be noted that the feedback gains are selected to minimize J that is given in Eq. (11.36). This means that the integrand of Eq. (11.36) must vanish as $t \to \infty$. Thus with a suitable choice of **Q**, the state $\mathbf{x}(t)$ will go to zero (which is the reference value for regulator action) with time and likewise for a suitable choice of **R**, the input $\mathbf{u}(t)$ will also go to zero for regulator action.

Linear Quadratic Regulator

Let the plant or system by defined by the following state and output equations:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{11.37a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{11.37b}$$

with $\mathbf{x}(t)$ as the $n \times 1$ state vector, $\mathbf{u}(t)$ as the $m \times 1$ control input and $\mathbf{y}(t)$ as the $p \times 1$ measured output. Referring to the previous chapter, the control law for full-state feedback is of the form

$$\mathbf{u} = -\mathbf{K}\mathbf{x} \tag{11.38a}$$

where **K** is an $m \times n$ matrix of constant coefficients called the controller gains. For the regulation action, if the controller **K** functions well for one specific constant reference, then it is as good for any other constant references. Therefore, for analysis and synthesis the reference is set to zero for regulator action without loss of generality. If one desires to feedback only the measurable output, then the control law for output feedback is of the form

$$\mathbf{u} = -\mathbf{K}\mathbf{y} \tag{11.38b}$$

where **K** is an $m \times p$ matrix of constant feedback coefficients called the controller gains that is to be determined. Since the control input vector $\mathbf{u}(t)$ is only allowed to depend on the output vector $\mathbf{y}(t)$ and not on the entire state vector $\mathbf{x}(t)$, this control law reflects the fact that only measurable quantities may be used to determine $\mathbf{u}(t)$. This implies that in output feedback, one is dealing with reduced state information.

The performance objective of state regulation for the system may be attained by selecting the control input $\mathbf{u}(t)$ to minimize a function J called the cost function or the performance index as given in Eq. (11.36).

Thus, for the linear quadratic regulator (LQR) problem, one can say that given a linear system as described by the model of Eqs. (11.37a) and (11.37b), the controller matrix **K** must be selected or designed for a regulator control input law as given in Eq. (11.38a) or Eq. (11.38b) such that the value of the performance index J as given in Eq. (11.36) is minimized. Such a controller is called the *linear quadratic regulator*.

By substituting the regulator control law of Eqs. (11.38a) and (11.38b) into the state equation [Eq. (11.37)], the closed-loop system is given by

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{BK})\mathbf{x} = \mathbf{A}_c\mathbf{x} \quad \text{(for full-state feedback)} \tag{11.39a}$$

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{BKC})\mathbf{x} = \mathbf{A}_c\mathbf{x} \quad \text{(for output feedback)} \tag{11.39b}$$

For any admissible choice of the performance index weighting matrices **Q** and **R**, closed-loop stability is guaranteed by this approach. Thus, **Q** and **R** may be treated as design parameters which are interactively varied until a suitable control gain **K** is obtained. One should contrast this with the approach used in classical techniques which might involve directly varying the elements of the control gain **K** in an attempt to obtain suitable closed-loop performance. In the classical approach, closed-loop stability is not even guaranteed for all values of **K**. Therefore it would be far more difficult to achieve an acceptable design. But here the design for tuning the gains in **K** for good performance where closed-loop stability may not even hold has been replaced by the problem of tuning the **Q** and **R** weighting matrices where closed-loop stability is at least guaranteed. Generally, the starting value of **Q** is chosen as $\mathbf{C}^T\mathbf{C}$ and **R** is chosen as identity matrix. The above discussions are valid even in the discrete domain. Without going into the derivations, the discrete LQR design algorithm is given in the next sub-section.

Linear Quadratic Regulator: Full-State Feedback

Let the given system or plant be expressed in the discrete domain as

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k \end{aligned} \tag{11.40}$$

For a full-state feedback control structure, the control law is given by

$$\mathbf{u}_k = -\mathbf{K}\mathbf{x}_k \tag{11.41}$$

The performance index or the cost function that is to be minimized is given as

$$J = \frac{1}{2}\sum_{k=0}^{\infty}(\mathbf{u}_k^T\mathbf{R}\mathbf{u}_k + \mathbf{x}_k^T\mathbf{Q}\mathbf{x}_k) \tag{11.42}$$

Note that the integrator is replaced by the discrete accumulator in the above cost function equation. Equations (11.36) and (11.42) are equivalent.

To find the controller gain matrix **K** that will minimize the cost function J as given in Eq. (11.42), the optimal solution is given by

$$\mathbf{K} = (\mathbf{R} + \mathbf{B}^T\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{P}\mathbf{A} \tag{11.43}$$

where **P** is an intermediate matrix variable which is obtained from the solution of

$$\mathbf{A}^T\mathbf{P}\mathbf{A} - \mathbf{P} + \mathbf{Q} - \mathbf{A}^T\mathbf{P}\mathbf{B}(\mathbf{R} + \mathbf{B}^T\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{P}\mathbf{A} = \mathbf{0} \tag{11.44}$$

Equation (11.44) is popularly known as the Discrete Algebraic Riccati Equation (DARE). The solution exists if **P** is positive definite. It may be cumbersome to find the solution to Eq. (11.44) manually. Therefore, one may have to use the help of software packages like MATLAB or SciLAB for solving Eq. (11.44). If DARE is a function that solves the DARE, one can find the **P** matrix using the command.

$$\mathbf{P} = \text{DARE}(\mathbf{A}, \mathbf{B}, \mathbf{Q}, \mathbf{R})$$

Equation (11.43) gives the LQR gain, **K** which is called the LQR gain and sometimes also called the *Kalman controller* gain.

Linear Quadratic Gaussian

If the controller is optimal and the estimator is a Kalman filter, then the combined system is known as linear quadratic Gaussian (LQG) system. In the LQG system, the estimation is optimal for Gaussian noise character. Here stability is always guaranteed with the optimal performance. Such a system is robust in performance also. In LQG scheme, both **K** and **G** are designed independently. **K** can be designed by LQR algorithm and **G** can be designed by Kalman filter approach. This approach has the advantages of

1. better noise rejection;
2. optimal control input;
3. optimal state estimate;
4. guaranteed stability.

Linear Quadratic Tracker: Output Feedback

The output feedback formulation was discussed in the previous chapter. This section discusses the performance index for the output feedback formulation given in Eqs. (10.101)–(10.103), reproduced here for easy reference:

$$x_{k+1} = Ax_k + Bu_k + Er_k$$

$$y_k = Cx_k + Fr_k$$

$$z_k = Hx_k$$

$$u_k = -Ky_k = -KCx_k - KFr_k$$

$$x_{k+1} = (A - BKC)x_k + (E - BKF)r_k = A_c x_k + B_c r_k$$

The input deviation is given by

$$\hat{u}_k = u_k - \bar{u}_k$$

$$\hat{u}_k = -KCx_k - KFr_k - (-KC\bar{x} - KF\bar{r}) = -KC\hat{x}_k \tag{11.45}$$

where $r_k = \bar{r} = r$ for a step reference. In the steady state, $x_{k+1} = x_k = x_\infty$, where x_∞ is the state under steady-state operating conditions given by

$$x_\infty = (I - A_c)^{-1} B_c r_\infty \tag{11.45a}$$

Here r_∞ is the reference input under steady-state operating conditions.

The steady-state error in response to a step of magnitude *r* is obtained by using the steady-state relationship of Eq. (11.45a). It is given by

$$\bar{e} = r - \bar{z} = r - H\bar{x} = [1 + H(A_c - I)^{-1} B_c] \tag{11.46}$$

The error deviation is given by

$$\hat{e}_k = e_k - \bar{e} = (r - Hx_k) - (r - H\bar{x}) = -H\hat{x}_k = -\hat{z} \tag{11.47}$$

where

$$\hat{z}_k = z_k - \bar{z} = Hx_k - H\bar{x} = H\hat{x}_k.$$

To make the tracking error, $e_k = \hat{e}_k + \bar{e}$ small, one may have to weight both $\hat{e}_k$ and $\bar{e}$ in a quadratic performance index. Thus the following performance index is used:

$$J = \frac{1}{2}\sum_{k=0}^{\infty}(\hat{e}_k^{\mathrm{T}} \cdot \hat{e}_k + \hat{u}_k^{\mathrm{T}} \cdot R \cdot \hat{u}_k) + \frac{1}{2}\bar{e}^{\mathrm{T}} \cdot V \cdot \bar{e} + \frac{1}{2}\sum_i \sum_j (g_{ij} \cdot k_{ij}^2) \tag{11.48}$$

Using Eq. (11.47), one can rewrite the performance index of Eq. (11.48) in terms of the state deviations as

$$J = \frac{1}{2}\sum_{k=0}^{\infty}(\hat{x}_k^{\mathrm{T}} \cdot Q \cdot \hat{x}_k + \hat{u}_k^{\mathrm{T}} \cdot R \cdot \hat{u}_k) + \frac{1}{2}\bar{e}^{\mathrm{T}} \cdot V \cdot \bar{e} + \frac{1}{2}\sum_i \sum_j (g_{ij} \cdot k_{ij}^2) \tag{11.49}$$

where $Q = qH^{\mathrm{T}}H$, $R > 0$ and $V \geq 0$ which is the steady-state error weighting matrix. The scalar weight q gives greater design flexibility in the choice of Q. It is generally acceptable to select $R = \rho I$ and $V = \mu I$, where ρ and μ are scalar design parameters. If the dynamic compensator in each loop contains a discrete integrator or a pole at $z = 1$, then the steady-state error will automatically be zero. In such cases, one can select $V = 0$ in the performance index function of Eq. (11.49). The usefulness of the above performance index is that steady-state errors can be guaranteed to be small even when some of the loops do not have integrators or poles at $z = 1$. The last term in the performance index weights the elements k_{ij} of the control gain matrix K. The motivation is to allow for more structure in the control system wherein experience and intuition can be used to suitably weight the elements of the control gain matrix K. Thus, if the error component of one loop does not influence the control input of another loop, the appropriate element of K may be made as small as desired by selecting its weight g_{ij} as a large value in the performance index. The first term in the performance index is the weighted quadratic state and input deviations which should become zero as time tends to infinity.

The question of concern to the control system designer is the selection of the weighting matrices Q and R. This will be discussed in the following section while illustrating the induction motor example. One must appreciate the fact that minimization of the performance index function as given in Eq. (11.49) may not always be the true design objective. The problem, however, is that the true design objective often cannot be expressed in mathematical terms and even when the design objective is expressible in mathematical terms, it may be impossible to solve for the optimum control law.

Optimal Gain for Output Feedback

The output feedback gains are obtained by solving a set of coupled equations. This section will focus on the design algorithm without going into the derivations for these coupled equations.

Design Equations

$$A_C^{\mathrm{T}} P A_C - P + Q + C^{\mathrm{T}} K^{\mathrm{T}} R K C = 0 \tag{11.50}$$

$$A_C S A_C^{\mathrm{T}} - S + X = 0 \tag{11.51}$$

$$RKCSC^{\mathrm{T}} - B^{\mathrm{T}}PA_{\mathrm{C}}SC^{\mathrm{T}} + B^{\mathrm{T}}(A_C - I)^{-\mathrm{T}}(P + H^{\mathrm{T}}VH)\bar{x}\cdot\bar{y}^{\mathrm{T}}$$
$$-B^{\mathrm{T}}(A_{\mathrm{C}} - I)^{-\mathrm{T}}H^{\mathrm{T}}Vr\bar{y}^{\mathrm{T}} + g * K = 0 \tag{11.52}$$

where r_k is a step of magnitude r and g^*K is a matrix with elements $g_{ij}k_{ij}$;

$$\bar{x} = -(A_{\mathrm{C}} - I)^{-1}B_{\mathrm{C}}; \quad \bar{y} = C\bar{x} + Fr; \quad X = \bar{x}\cdot\bar{x}^{\mathrm{T}}$$

$$A_{\mathrm{C}} = A - BKC$$

$$B_{\mathrm{C}} = E - BKF$$

P is a positive definite constant matrix, that is, $P > 0$.

Performance Index Constraint The performance index given in Eq. (11.49) is the cost that needs to be minimized. This performance index can be shown to be equivalent to the following function [given in Eq. (11.53)]. The first term of Eq. (11.49) is replaced with an easily implementable equivalent given as the *trace of PX* as indicated in Eq. (11.53).

$$J = \frac{1}{2}\mathrm{tr}(PX) + \frac{1}{2}\bar{e}^{\mathrm{T}}\cdot V\cdot\bar{e} + \frac{1}{2}\sum_i\sum_j(g_{ij}\cdot k_{ij}^2) \tag{11.53}$$

where $\bar{e} = [I + H(A_{\mathrm{C}} - I)^{-1}B_{\mathrm{C}}]r$

One should note that the optimal feedback gain K is obtained by solving the set of coupled equations [Eqs. (11.50)–(11.52)] by minimizing the performance index given in Eq. (11.53). This minimization problem may be solved by using either gradient-based algorithms or non-gradient-based algorithms like the SIMPLEX algorithm. Many popular software packages provide subroutines for minimization algorithms. If the SIMPLEX algorithm is used, then the optimal feedback gain K can be obtained by using only Eq. (11.50) which is also called the Lyapunov equation. However, for gradient-based routines which are faster than the SIMPLEX approach, all three equations [Eqs. (11.50)–(11.52)] are required. If the system is controllable and observable, and the Q and R matrices are appropriately chosen, the minimization of J will not fail to find a K which stabilizes the closed-loop system. This means that once the system along with Q and R is specified, determination of K is a problem in numerical analysis. This does not mean that minimizing J is a trivial problem. There are many standard algorithms used to find the global minimum. Two of the standard methods used that are popular for finding the global minimum are:

1. Find the local minimum starting from widely varying starting values of the independent variables and then pick the minimum among these.
2. Perturb a local minimum by taking a finite amplitude step away from it and then see if the routine returns to a better point or always to the same one.

11.8 Induction Motor Example

The induction motor has been formulated as an LQ tracker problem with output feedback in Section 10.14. The objective now is to determine the gain matrix K for the controller structure (which is a proportional–integral structure in this case) so as to minimize a specified performance index. From the performance index J, as given in Eq. (11.53), one can set $V = 0$ in this specific case, because the steady-state error is zero due to the use of the proportional–integral structure which has a pole at $z = 1$.

From the gain matrix given in Eq. (10.116) (Chapter 10), it is clearly seen that some of the terms in the K matrix are zero. Therefore, in the performance index function of Eq. (11.53), the corresponding matrix

elements k_{ij} can be weighted with a very large value of g_{ij}, so that after minimization the resulting gain values of these k_{ij} elements will be very small and can be set to zero during implementation. This weighting of the elements of K has to be done for minimization algorithms using the gradient approach. For the SIMPLEX approach, those matrix elements which are zero can be set to zero and the remaining matrix elements can be optimally obtained by minimization of the performance index given by

$$J = \frac{1}{2}\,\mathrm{tr}(PX) \tag{11.54}$$

which is obtained from Eq. (11.53). As a consequence, it would imply that with SIMPLEX approach, one can fix any gain element of the gain matrix K and obtain the optimal gain values for the rest of the elements of the gain matrix K by minimizing the performance index J, as given in Eq. (11.54). This in fact gives lots of flexibility in arriving at stable optimal gains for the proportional–integral control structure of the system.

The steps for obtaining the optimal gain values of the gain matrix K, once the system is formulated, as given in Eqs. (10.101a)–(10.102), is as follows:

Step 1: The optimization process is an iterative process. Therefore, to start the optimization process, an initial gain matrix K, which makes the closed-loop plant (i.e., $A - BKC$) stable, has to be used. Such a gain may or may not be easy to find. One reliable though tedious way to find a stabilizing gain is to use discrete root locus techniques. The gain can then be optimized using the optimization of the performance index J. However, as the induction motor is open-loop stable which is established from the eigenvalues of the induction motor system matrix, the compensator and the induction motor system combination is consequently open-loop stable. Therefore, all the open-loop poles lie within the unit circle. If the gains are very small, then the closed-loop poles are very close to the open-loop poles and hence, the closed-loop system is stable. As a consequence, the proportional gains can be set to zero and very small values of integral gains (around 0.01) can be used. This will usually provide the starting eigenvalues of $A - BKC$, such that they are within the unit circle.

Step 2: The state weighting matrix Q and the input weighting matrix R need to be chosen. From Eq. (11.49) one obtains the state weighting matrix Q, which is equal to $H^{\mathrm{T}}H$. However, to tune the response, a scalar weight q is used such that

$$Q = q \cdot H^T H$$

Regarding the input weighting matrix R, it is usual to choose $R = \rho I$, where ρ is a scalar. If a large value of ρ is used, the resulting gain K leads to small inputs to the system, thereby slowing down the system response. If one needs to have a slower or faster response of the system with respect to some input, then the corresponding element in the R matrix can be suitably weighted higher or lower, respectively.

Step 3: The performance index J in Eq. (11.54) is minimized under the constraint of the Lyapunov equation [Eq. (11.50)] using the SIMPLEX algorithm.

In the case of the induction motor, it is evident from Eq. (10.104) that the system matrix is dependent on the rotor speed ω_m. As a consequence, one would obtain different optimal gains at different rotor speeds. The plot of the K_{pd}, K_{id}, K_{pq} and K_{iq} versus the rotor speed ω_m is shown in Figures 11.13(a) and (b). The gains are calculated for eight different speeds as shown in Figure 11.13. For other speed values, the optimum values of the gains are obtained by interpolation. A lookup table can be used in the digital controller of the induction motor drive system such that the appropriate optimal gain values are used at a specific speed. On the other hand, one may also use a fixed gain matrix K corresponding to the maximum rotor

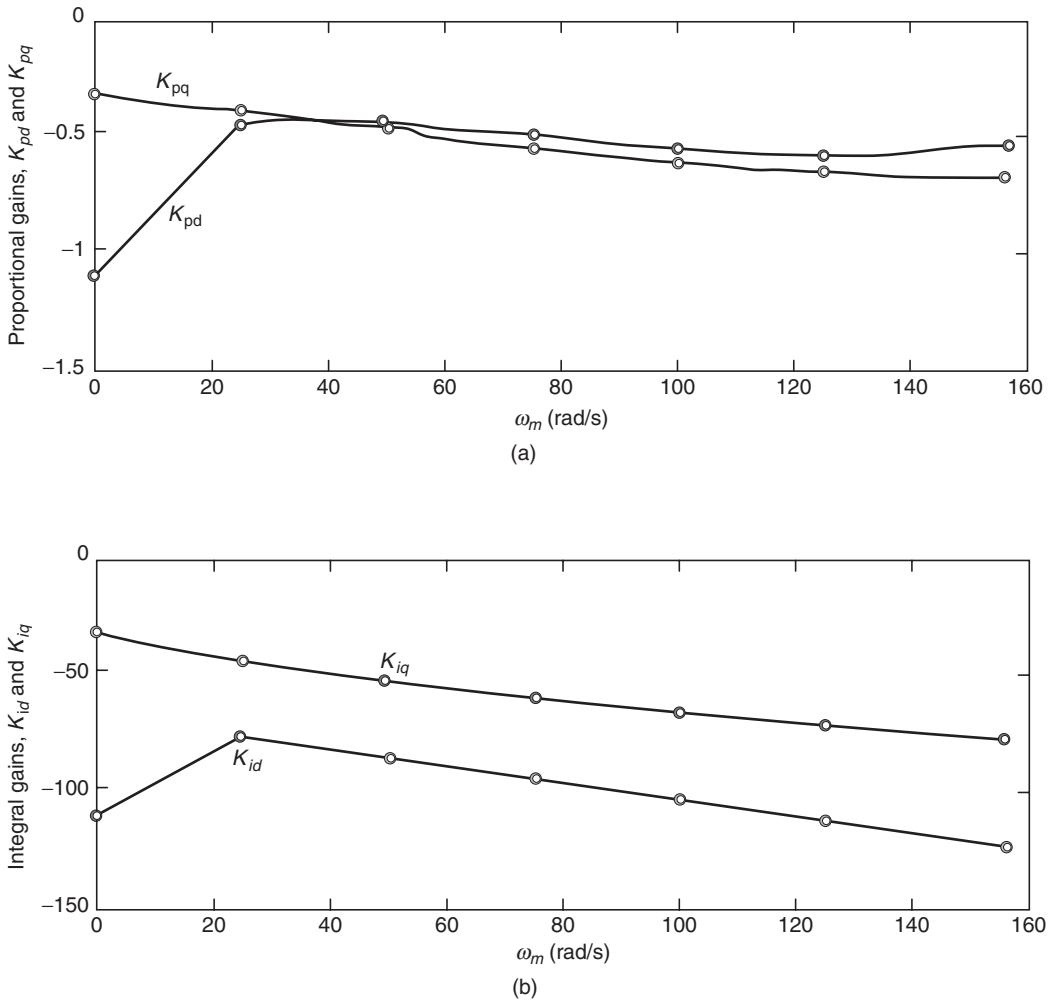


Figure 11.13 (a) Optimal proportional gains as a function of shaft speed; (b) optimal integral gains as a function of shaft speed.

speed. This would lead to sub-optimal operation at other speed values. But nonetheless, stability is guaranteed at all speed ranges. The induction motor used for controller design has the parameters indicated in Appendix V.

As lots of analytical and experimental experience is available on the induction motor, the designer should put the intuition and experience to advantage while using the output-feedback approach. In fact, the proportional–integral controller structure that was proposed while formulating the induction motor drive system as an LQ tracker problem is based on the experience of induction motor operation in constant

flux operation schemes such as the *V*/*f* controls. It is evident that in this context, the output feedback is appropriate and in view of the highly coupled dynamics of the current controllers, multi-variable design techniques discussed till now become more or less necessary.

It is discussed in the previous chapter (sub-section "Vector Control" of Section 10.12) that good dynamic performance is obtained only under field orientation. With the orientation of the arbitrary axis along the rotor field and the use of the decoupling network, the induction motor plant as seen by each current control loop reduces to a first-order system. However, when the rotor field is disoriented, then the induction motor is a fourth-order system wherein the dynamic performance deteriorates compared to dynamic performance under rotor field orientation. This is explicitly indicated in Figure 11.14 for the i_{sd} and i_{sq} currents.

The performance index J in Eq. (11.53) is also very flexible, which is conducive for intuitive manipulation of the weights of the gains g_{ij}. The following strategy has been found useful in the context of the induction motor for arriving at the optimal gains.

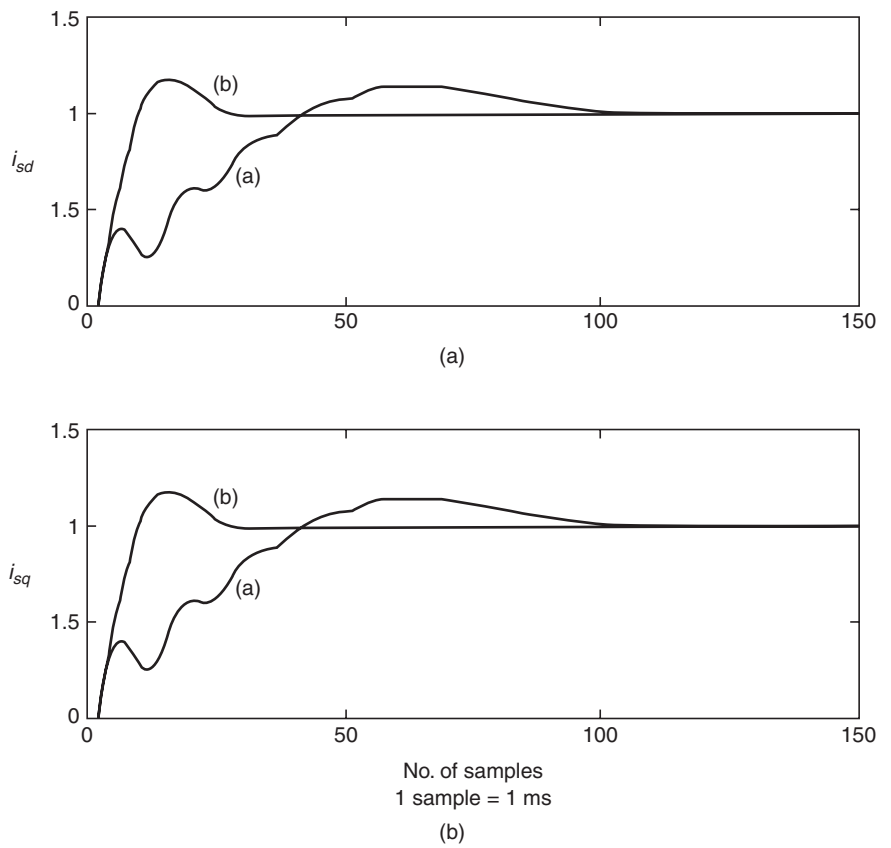


Figure 11.14 Step response for i_{sd} and i_{sq}: (a) Arbitrary axis is not aligned along the rotor field; (b) arbitrary axis is aligned along the rotor field, that is, rotor field-orientated control. ($Q = 0.1H^{T}H$, R = diag[1, 20], K_{gains} = [–0.3, –62.1088, –0.3, –48.572].)

1. Start with the initial gain values of $K_{gains} = [K_{pd}, K_{id}, K_{pq}, K_{iq}] = [0, -0.01, 0, -0.01]$.
2. Set $Q = H^T H$ and $R = \text{diag}(1,1)$ where the first element of the diagonal of R corresponds to the direct axis (d-axis) input and the second element of the diagonal of R corresponds to the quadrature axis (q-axis) input.
3. Obtain the optimal gain matrix K by solving Eq. (11.50) (by SIMPLEX minimization method) under the constraint of Eq. (11.53).
4. Check the step response of the closed-loop system.
5. If the step response is not proper, change the input weighting matrix R and the state weighting matrix Q and obtain K till proper response is obtained.
6. If one has to slow down the system, decrease Q and/or increase R. If one has to speed up the system then increase Q and/or decrease R, as shown in Figure 11.15.

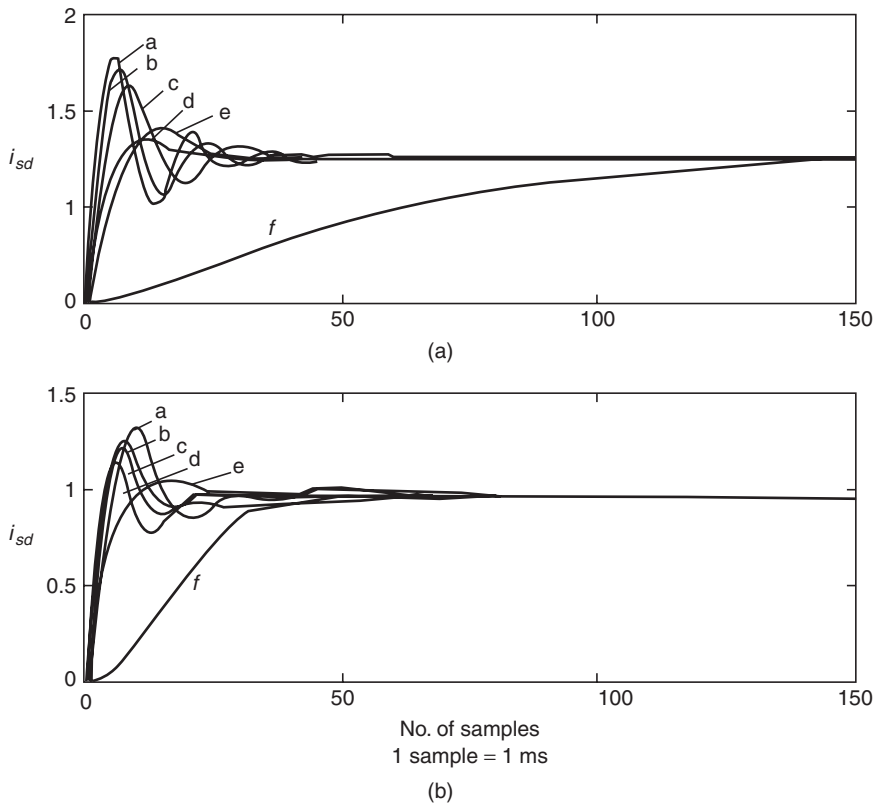


Figure 11.15 Step response of i_{sd} and i_{sq} under rotor field-oriented control where a, b, c, d, e, f responses are for the following gains, Q and R. $K_{gains} = [K_{pd}\ K_{id}\ K_{pq}\ K_{iq}]$; Q; R. (a) [–0.9404, –146.1699, –0.4600, –99.7845]; $Q = 10H^TH$; $R = \text{diag}(0.1, 2)$.(b) [–0.7785, –141.9508, –0.5872, –89.7422]; $Q = H^TH$; $R = \text{diag}(0.1, 2)$. (c) [–0.5457, –124.2175, –0.6775, –77.3730]; $Q = 0.1H^TH$; $R = \text{diag}(0.1, 2)$. (d) [–0.4231, –62.1088*, –0.7961, –47.9323]; $Q = 0.1H^TH$; $R = \text{diag}(1, 20)$. (e) [–0.3*, –62.1088*, –0.3*, –48.5720]; $Q = 0.1H^TH$; $R = \text{diag}(1, 20)$. (f) [–0.0001*, –5*, –0.0001*, –15.6961]; $Q = H^TH$; $R = \text{diag}(1, 20)$.

Note: *These gains were fixed during optimization (constrained optimization).

7. If one has to speed up or slow down the system controller to a large extent, one could adopt the following method which allows more structure to the control by way of increased intuitive interaction by the designer. For example, if one has to slow down the response, fix the proportional gain values K_{pd} and K_{pq} to a small value. Now set the input weight matrix R such that $R = \text{diag}(rwd, 1)$ where rwd is a very large value. While optimizing, if gradient algorithms are used, weight the K_{pd} and K_{pq} gains with large weights. If the simplex algorithm is used, just fix the values of K_{pd} and K_{pq} to low values. After optimization, the optimal K_{id} value obtained will be a small value. Now set $R = \text{diag}(1, rwq)$ where rwq is a large weight on the q-input. Fix the K_{id} gain to the value obtained during the previous optimizing run. Now the gain value K_{iq} obtained will also be low. K_{id} and K_{iq} can be iteratively fixed and optimized to obtain any degree of slowness of speed of response.

Referring to Figure 11.15, it is evident that as Q increases, the response becomes faster. One should also note that the speed of response of i_{sd} and i_{sq} need not be identical. This is because under rotor field-oriented control, the i_{sd} and i_{sq} components are decoupled. From Figure 11.15, it is evident that response (e) is preferable from the point of view of performance criteria. However, it will be shown in the next section that response (f) is most appropriate because the controller gains corresponding to the response (f) give the best stability robustness compared to the other gain values. More importantly, under rotor field-oriented control, upto base speed, the i_{sd} component is kept constant and the torque dynamics is determined only by the i_{sq} component. Therefore, it is sufficient that i_{sq} has good dynamic performance. In the next section, it will be shown that the gains corresponding to response (e) can be used if the rotor resistance adaptation is performed to improve stability robustness. In the following section, robustness concepts are considered along with the design methodology using the induction motor example for illustration.

11.9 Robust Controller Design

In the previous section it is assumed that an exact state-variable description of the plant to be controlled is available. However, in a practical situation, the actual plant and the model plant will never be identical. Some of the parameters of the system matrix may be different from that of the actual system. In the case of the induction motor example, parameters like the rotor resistance, the stator resistance or the magnetizing inductance of the actual motor may be at variance from the corresponding parameters of the motor model, which could lead to deterioration in performance and stability. The design of the current controllers considered in the context of robust controller design for induction motor accommodates for the uncertainty in the system parameters.

Further, it is often important to account for disturbances and sensor measurement noises. Disturbances may act to cause unsatisfactory performance in a system. Therefore, it is important to design controllers that have stability robustness and performance robustness. Stability robustness is the ability to guarantee closed-loop stability in spite of parameter variations. Performance robustness is the ability to guarantee acceptable performance even though the system may be subject to disturbances.

Consider the standard closed-loop system shown in Figure 11.16. A few signals have been added to characterize the uncertainties. The signal $d(t)$ represents a disturbance acting on the system and $n(t)$ represents the sensor or measurement noise. If $S(z)$ is the system sensitivity and $T(z)$ is the system co-sensitivity, then referring to Figure 11.16, they are defined as

$$S(z) = (I + G_{\text{pa}}\, G_{\text{c}})^{-1} \tag{11.55}$$

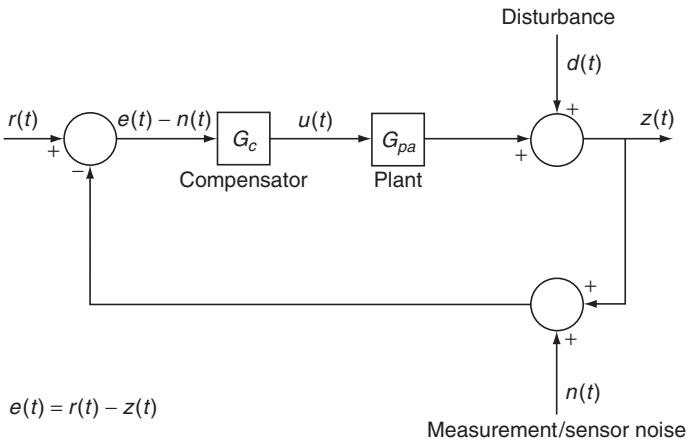


Figure 11.16 | Standard closed-loop system with disturbance input and measurement noise.

$$T(z) = (I + G_{pa}G_c)^{-1} \cdot G_{pa}G_c = G_{pa}G_c(I + G_{pa}G_c)^{-1} \tag{11.56}$$

As $S(z) + T(z) = I$, $T(z)$ is called the *complementary sensitivity* or the *co-sensitivity*.

Referring to Figure 11.16, it can be shown that the performance output $z(t)$ and the tracking error $e(t)$ are given in the z-domain by

$$Z(z) = T(z)[R(z) - N(z)] + S(z)D(z) \tag{11.57}$$

$$E(z) = S(z)[R(z) - D(z)] + T(z)N(z) \tag{11.58}$$

The disturbance $d(t)$ occurs usually at low frequencies and the measurement noise $n(t)$ is predominant at high frequencies. On examining Eqs. (11.57) and (11.58), it is observed that at low frequencies where the disturbances are predominant, the sensitivity S must be small, and at high frequencies where the measurement noises are predominant, the co-sensitivity T must be small. To obtain a measure of the "smallness" of the sensitivity S and the co-sensitivity T, multi-variable Bode plots are used.

The multi-variable Bode magnitude plot is nothing but the plot of the transfer function singular values versus the frequency. MIMO generalizations of the Bode magnitude and phase plots have been reported in the literature. However, the theory relating to the phase portion of the plot is more difficult to use in a practical design technique. In the case of robust control design for a discrete system, the magnitude plot is often sufficient. If $H(j\omega)$ is a square transfer function matrix (which is true in our case) at any frequency ω, then *for any given input, the magnitude of the transfer function $H(j\omega)$ at any given frequency ω is bounded above by its maximum singular value, denoted by $\bar{\sigma}(H(j\omega))$ and bounded below by its minimum singular value, denoted by $\underline{\sigma}(H(j\omega))$*. Therefore for design of the robust controller, one needs to consider only these two constraining values of the magnitude.

When performing algebra with singular values, the following relations with respect to singular values are useful:

For any matrices A, B, GK and non-singular matrix M.

$$\underline{\sigma}(GK) - 1 \le \underline{\sigma}(I + GK) \le \underline{\sigma}(GK) + 1 \tag{11.59}$$

$$\bar{\sigma}(M) = \frac{1}{\underline{\sigma}(M^{-1})} \tag{11.60}$$

$$\bar{\sigma}(AB) \le \bar{\sigma}(A)\bar{\sigma}(B) \tag{11.61}$$

Frequency Domain Performance Specifications

It is desired that $S(j\omega)$ be small at low frequencies and $T(j\omega)$ be small at high frequencies. From Eq. (11.55) and using relation in Eq. (11.60), one obtains

$$\bar{\sigma}(S) = \bar{\sigma}[(1 + G_{pa}G_c)^{-1}] = 1/\underline{\sigma}(I + G_{pa}G_c) \approx 1/\underline{\sigma}(G_{pa}G_c) \tag{11.62}$$

From Eq. (11.56) and using relation in Eq. (11.60), one obtains

$$\bar{\sigma}(T) = \bar{\sigma}[G_{pa}G_c(I + G_{pa}G_c)^{-1}] \approx \bar{\sigma}(G_{pa}G_c) \tag{11.63}$$

From Eq. (11.62) it is evident that a large value of $\underline{\sigma}(G_{pa}G_c)$ guarantees a small value of $\bar{\sigma}(S)$. From Eq. (11.63), it is evident that a small value of $\bar{\sigma}(G_{pa}G_c)$ will guarantee a small value of $\bar{\sigma}(T)$. Therefore, the specifications that $S(j\omega)$ should be small at low frequencies and $T(j\omega)$ should be small at high frequencies may well be formulated in terms of the *minimum singular value*, $\underline{\sigma}(G_{pa}G_c)$ *being large at low frequencies* and the *maximum singular value* $\bar{\sigma}(G_{pa}G_c)$ *being small at high frequencies*. Thus, all performance specifications can be in terms of the minimum and maximum singular values of the loop gain, $G_{pa}G_c$. The practical significance is that one needs to examine only the singular values and not $S(j\omega)$ and $T(j\omega)$. Further, software packages give standard routines to compute the singular values of any transfer function. This makes the robust controller design much simpler.

The sensor or measurement noise is generally appreciable at high frequencies. Therefore, to keep the tracking error small in the face of the measurement noise, the co-sensitivity, $T(j\omega)$ should be small at high frequencies above a certain frequency, say ω_n. This may be guaranteed if

$$\bar{\sigma}(G_{pa}G_c) \ll 1, \text{ for } \omega \ge \omega_n \tag{11.64}$$

Modeling Uncertainties

To guarantee *stability robustness* in the face of plant modeling uncertainty, the uncertainties may be included into the model in one of the following two ways: (a) additive uncertainties or (b) multiplicative uncertainties.

The actual plant model G_{pa}^* and the assumed plant model G_{pa} may differ by additive uncertainties such that

$$G_{pa}^* = G_{pa} + \Delta G_{pa} \tag{11.65}$$

where the singular value of the additive uncertainty satisfies a known bound

$$\bar{\sigma}(\Delta G_{pa}) < a(\omega) \tag{11.66}$$

with $a(\omega)$ known for all ω. On the other hand, the actual plant model G_{pa}^* and the assumed plant model G_{pa} may differ by multiplicative uncertainties, such that

$$G_{pa}^* = [I + M(j\omega)]G_{pa} \tag{11.67}$$

where the singular value of the multiplicative uncertainty satisfies a known bound

$$\bar{\sigma}[M(j\omega)] < m(\omega) \tag{11.68}$$

with $m(\omega)$ known for all ω. Since Eq. (11.65) can be written as

$$G_{pa}^{*} = [I + \Delta G_{pa} G_{pa}^{-1}] G_{pa} = [1 + M(j\omega)] G_{pa} \tag{11.69}$$

one can confine the discussion to only multiplicative uncertainties. For stability robustness, the co-sensitivity $T(j\omega)$ should be bounded above by the reciprocal of the multiplicative modeling discrepancy bound $m(\omega)$ which is given by

$$\bar{\sigma}[G_{pa} G_c (I + G_{pa} G_c)^{-1}] = \bar{\sigma}(T) < 1/m(\omega) \tag{11.70}$$

If the worst case uncertainties in the system parameters are known, then the upper bound $m(\omega)$ can be found from Eq. (11.67) as follows:

$$M(j\omega) = (G_{pa}^{*} - G_{pa}) G_{pa}^{-1} \tag{11.71}$$

$$m(\omega) = \bar{\sigma}[M(j\omega)] \tag{11.72}$$

The robustness bound $1/m(\omega)$ and the maximum singular values of $T(j\omega)$ for the system structure of Figure 11.16 are plotted for an induction motor plant by using a computer program (Figure 11.17).

In Figure 11.17, the robustness bound $1/m(\omega)$, for rotor resistance, stator resistance and magnetizing inductance parameter uncertainties (from 33% to 99%) are shown. In the case of the rotor and stator resistance uncertainties as the actual parameters increase from the nominal values (which is the cold resistance value used in the assumed model), the upper bound for $T(j\omega)$ decreases. Though the magnetizing inductance parameter uncertainty is shown upto 100% deviation from the nominal, in practice, the uncertainty in the magnetizing inductance parameter will not be more than 10–20%. One can observe from Figure 11.17 that uncertainties in the magnetizing inductance will also lead to lowering the permissible system bandwidth.

Controller Design Steps

The algorithm for robust controller design for the induction motor drive system which is formulated as an LQ tracker problem with output feedback is as follows.

Step 1: Augment the entire drive system as an LQ tracker problem as discussed in Section 10.14 in the previous chapter.

Step 2: Proper choices for Q and R matrices are made and the optimal gain matrix K is obtained using either SIMPLEX minimization or gradient descent minimization techniques.

Step 3: Simulate the time responses of the closed-loop system to verify that they are satisfactory. If the responses are not proper, go to Step (2).

Step 4: Determine the upper bound for stability robustness. Plot the maximum singular values of $T(j\omega)$ to verify that it is within the upper bound $1/m(\omega)$. If it is not satisfied, return to Step (2).

The robust stability bound and the maximum singular values of $T(j\omega)$ for the response (e) of Figure 11.15 are plotted in Figure 11.18(a). It is seen that the upper bound for guaranteed robust stability is not satisfied as the $\bar{\sigma}(T)$ goes beyond the upper bound at certain frequencies. Therefore, to guarantee robust stability one should reduce the controller bandwidth. In Figure 11.18(b), the robust stability bound and the maximum singular values of $T(j\omega)$ for the response (f) of Figure 11.18 is plotted. It is evident that the gains corresponding to the response (f) of Figure 11.18 guarantees stability robustness. One should note that stability robustness and speed of response are generally contradictory specifications. Hence, it turns out

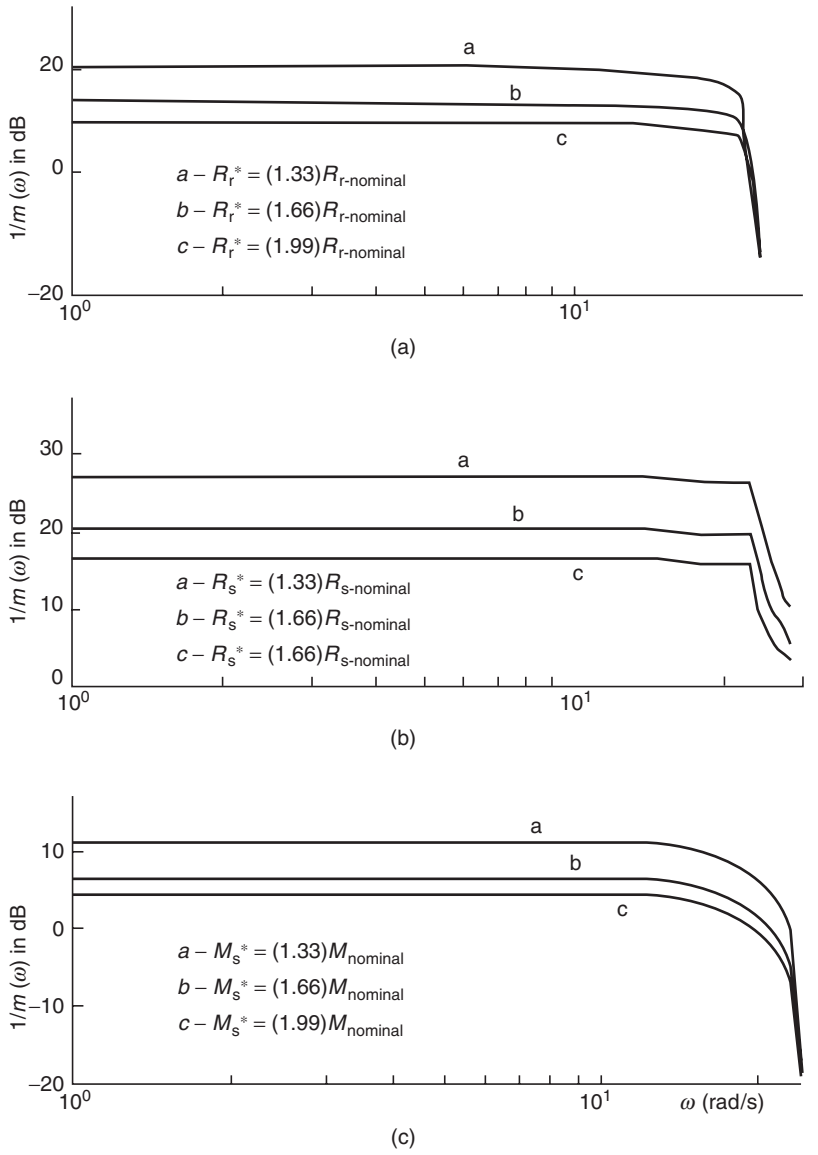


Figure 11.17 $1/m(\omega)$ (robustness bounds) for: (a) Rotor resistance uncertainties; (b) stator resistance uncertainties; (c) magnetizing inductance uncertainties, at $\omega_m = 157$ rad/s.

as a classic case of compromise. It is evident that as the controller bandwidth is reduced, the system becomes more and more robust. This same argument suggests that field-oriented control of an induction motor provides less stability robustness than scalar flux control techniques such as the *V*/*f* control strategies. However, field-oriented control scores over *V*/*f* control when high dynamic performance is required.

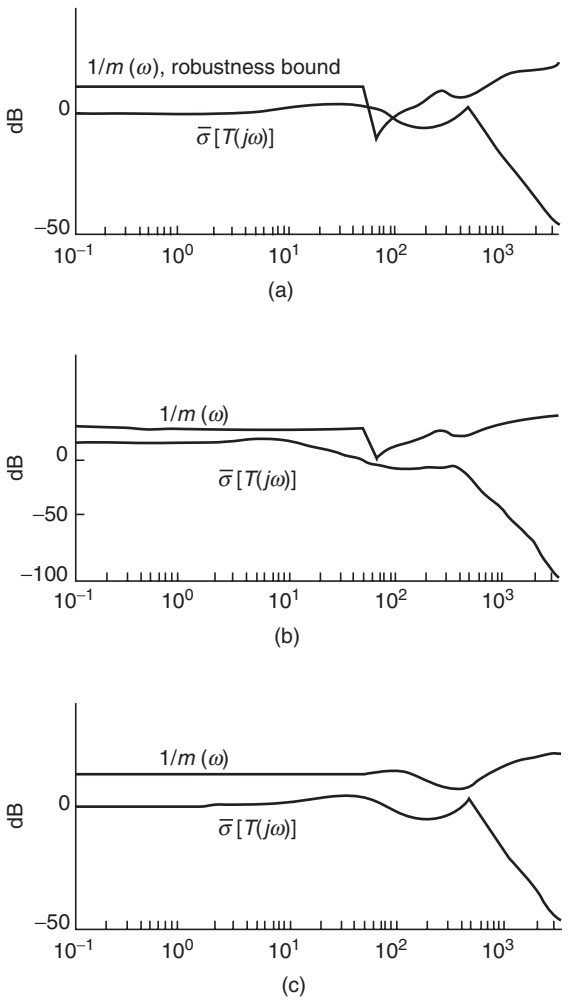


Figure 11.18 (a) Stability robustness with 100% uncertainty in rotor resistance, 100% uncertainty in stator resistance and 10% uncertainty in magnetizing inductance $\{Q = 0.1H^{T}H$, $R = \text{diag}(1, 20)$, $K_{\text{gains}} = [-0.3, -62.1088, -0.3, -48.572]\}$. (b) Stability robustness with reduced bandwidth, but with 100% uncertainty in rotor resistance, 100% uncertainty in stator resistance and 10% uncertainty in magnetizing inductance $\{Q = H^{T}H$, $R = \text{diag}(1, 20)$, $K_{\text{gains}} = [-0.0001, -5, -0.0001, -15.6961]\}$. (c) Stability robustness when the rotor resistance is adapted within 20% of the actual value even though there is 100% uncertainty in stator resistance and 10% uncertainty in magnetizing inductance $\{Q = 0.1H^{T}H$, $R = \text{diag}(1, 20)$, $K_{\text{gains}} = [-0.3, -62.1088, -0.3, -48.572]\}$.

An alternative to reducing the bandwidth of the controller to improve robustness would be to adapt the parameters of the assumed model to that of the actual model. In Figure 11.18(c), it is shown that the stability robustness of the gains corresponding to the response (*e*) of Figure 11.18 has been improved by adapting the rotor resistance to within 20% of the actual value, that is, only partially adapted. In the case of the induction motor plant with uncertainties in the rotor resistance, stator resistance and the magnetizing inductance, it is evident from Figure 11.18(c) that the stability robustness property of the system improves when the rotor resistance of the assumed model is adapted towards the actual plant parameter value. This is because the system is more sensitive to uncertainties in the rotor resistance than the stator resistance. Therefore, adaptation of the rotor resistance parameter improves not only the field orientation, which enhances the dynamic performance, but also improves stability robustness.

Speed Controller Synthesis

The current control loop of the induction motor drive system is very fast compared to the speed control loop. This is because the time constants of the stator currents are much faster than that of the mechanical time constant of the system. Therefore, with respect to the speed loop, the plant can be approximated as a first-order system with a lag corresponding to the mechanical time constant of the rotor. The control block diagram for the speed control system is shown in Figure 11.19. The anti-alias filter is not shown in this case because normally the speed is sensed using encoders which are directly compatible with digital processors.

As this is a first-order system, one could adopt standard classical SISO techniques, like the root locus or the Bode diagram techniques to design the speed controller. But one cannot get optimal gain values. Therefore, using the same framework as discussed for the digital current controller synthesis, the speed control model is also formulated as a LQ tracker problem with output feedback.

From Eq. (8.125), the plant $G_{p\omega}$ dynamics is described by matrices $A_{p\omega}$, $B_{p\omega}$, $C_{p\omega}$ and $D_{p\omega}$, where

$$A_{p\omega} = \left[-\frac{B}{J}\right] \tag{11.73}$$

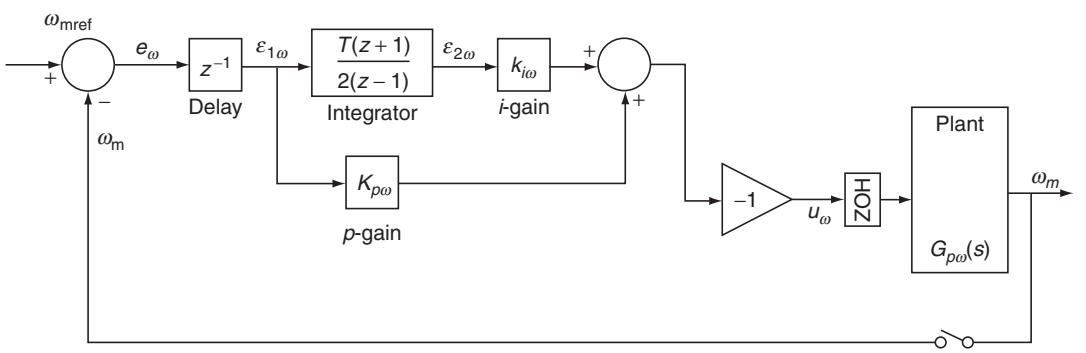


Figure 11.19 | Control block schematic of the plant for the speed controller design.

(B is the friction coefficient; J the mechanical inertia, referred to the rotor shaft)

$$B_{\mathrm{p}\omega} = \left[\frac{1}{J}\right] \tag{11.74}$$

where the state of $G_{p\omega}$ is given by

$$x_{\mathrm{p}\omega} = \left[\frac{\omega_{\mathrm{m}}}{p}\right]^{\mathrm{T}}$$

the input of $G_{p\omega}$ is given by

$$u_{\mathrm{p}\omega} = [u_{\omega}]^{\mathrm{T}}$$

the output of $G_{p\omega}$ is given by

$$y_{\mathrm{p}\omega} = x_{\mathrm{p}\omega}$$

This is a special case of output feedback which is the full-state feedback. With full-state feedback, the LQ tracker problem is much simpler and also the robustness and stability properties are better. However, following the same underlying framework laid down for the current controller synthesis, the dynamics of the compensator here is also a proportional–integral structure. One can then describe the controller $G_{c\omega}$ by $A_{c\omega}$, $B_{c\omega}$, $C_{c\omega}$ and $D_{c\omega}$ where

$$A_{\mathrm{c}\omega} = \begin{bmatrix} 0 & 0 \\ \dfrac{T}{2} & 1 \end{bmatrix}^{\mathrm{T}} \tag{11.75}$$

(where T is the sampling interval in the discrete controller);

$$B_{\mathrm{c}\omega} = \begin{bmatrix} -1 & 1 \\ -\dfrac{T}{2} & \dfrac{T}{2} \end{bmatrix} \tag{11.76}$$

where the state of $G_{c\omega}$, $x_{c\omega}$ is

$$x_{\mathrm{c}\omega} = [\varepsilon_{1\omega}, \varepsilon_{2\omega}]^{\mathrm{T}}$$

the input of $G_{c\omega}$, $u_{c\omega}$ is

$$u_{\mathrm{c}\omega} = \left[\frac{\omega_{\mathrm{m}}}{p}, \frac{\omega_{\mathrm{mref}}}{p}\right]^{\mathrm{T}}$$

the output of $G_{c\omega}$, $y_{c\omega}$ is

$$y_{\mathrm{c}\omega} = x_{\mathrm{c}\omega}$$

The plant $G_{p\omega}$ and the compensator $G_{c\omega}$ are combined so that the speed control system is formulated as LQ tracker problem for the augmented system given by Eqs. (10.101)–(10.102). The same steps used for the

current controllers are also used for the speed controller design. Figure 11.20 shows the step responses of the shaft speed for various values of the state weighting matrix Q. It is seen that as Q increases, the speed of response increases. The stability robustness for the speed control system for 100% uncertainty in parameters like J and B is shown in Figure 11.21 for the gains of response (a) of Figure 11.20. It can be seen that the system can withstand large uncertainties in parameters J and B, without affecting the stability robustness.

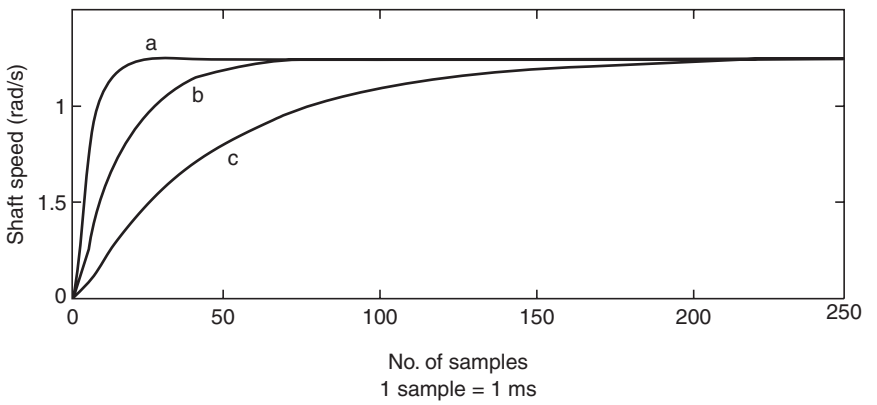


Figure 11.20 Step response for shaft speed: (a) $Q = 100H^TH$, $R = 0.2$, $[K_{p\omega}, K_{i\omega}] = [-16.9404, -1.6134]$; (b) $Q = 10H^TH$, $R = 0.2$, $[K_{p\omega}, K_{i\omega}] = [-6.411, -0.6356]$; (c) $Q = H^TH$, $R = 0.2$, $[K_{p\omega}, K_{i\omega}] = [-2.1645, -0.2163]$.

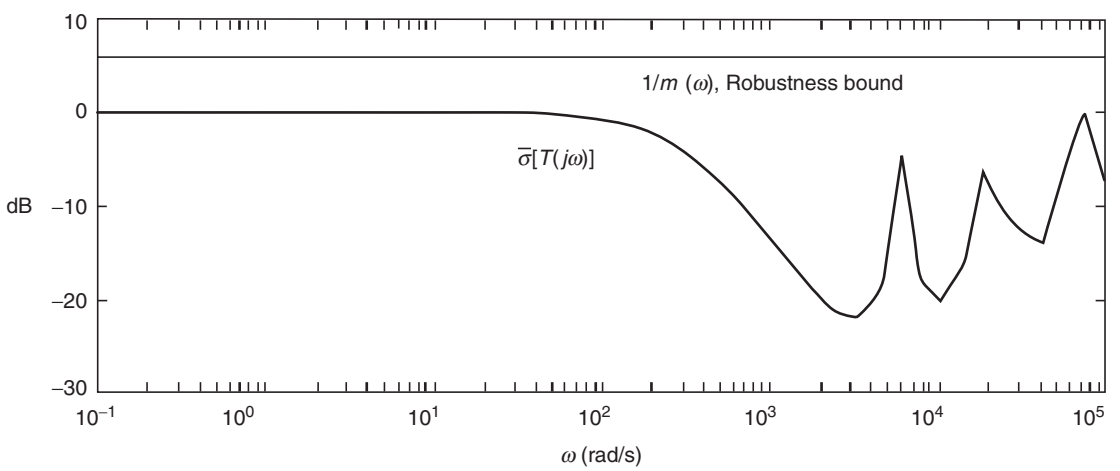


Figure 11.21 Stability robustness with 100% uncertainty in J and 100% uncertainty in $Q = 100H^TH$, $R = 0.2$, $[K_{p\omega}, K_{i\omega}] = [-16.9404, -1.6134]$.

CONCLUDING REMARKS

This chapter discusses some advanced aspects of controller design that addresses optimality issues and robustness issues. These controller design concepts are generally employed for complex and high performance systems. However, building optimality and robustness into the controller and estimator designs is an added feature that one may incorporate to achieve better energy efficiencies, stability and performance quality.

It must be understood that higher the desired speed of response, higher will be the instantaneous power that will be drawn from the power source. Optimal controllers will not change this principle, but they would ensure that for a given speed of response, the power drawn will be minimum if the energy constraints are built into the cost function or performance index. Optimal controllers generally operate on minimizing the state and input energy. On the other hand, the optimal estimators operate on minimizing the state error, though the generic problem of minimization of a cost function holds in both cases.

Most systems will have their system parameters varying during operation. This may be due to temperature drifts, change in operating points or heating due to load conditions. As a consequence the parameters of the actual system are at variance with respect to the system parameters used in the controller design. If the closed-loop pole locations are close to the stability boundary, then there is a possibility that the system may drift into the unstable zone or the performance may deteriorate. Robust controllers address these issues. In the face of uncertainties, in parameters and disturbances, robust controllers ensure that stability and performance are not compromised. This would actually result in controllers where the closed-loop system poles are located close to the origin of the unit circle. In that sense, the dead beat controller is a very robust controller.

TUTORIAL EXERCISES

The following exercises can be studied using MATLAB.

1. A first-order system is given. This is in the form of a file sys_mod.p which is a MATLAB file. The file sys_mod.p is included in the accompanying CD. The system has the following syntax:

$$y = \text{sys\_mod}(u)$$

where u is the input and y is the output. The model of the system given in sys_mod.p is of the form

$$(1 + b_1 z^{-1}) y_k = (a_0 + a_1 z^{-1}) u_{k-1}$$

The objective is to estimate a_0, a_1 and b_1 using the least square algorithm for minimizing the error. This involves the following tasks:

   a. Write a general purpose **least square algorithm** as an *m*-function.
   b. Formulate the problem for least square solution.
   c. Generate input and output data for the system (recommendation – use random inputs).
   d. Estimate the parameters.
   e. Plot the evolution of error and parameters with time.

   After trying out the algorithm on the above system, check if the algorithm works for another system which is of the form

$$(1 + b_1 z^{-1} + b_2 z^{-2}) y_k = (a_0 + a_1 z^{-1} + a_2 z^{-2}) u_{k-1}$$

   wherein one has to estimate a_0, a_1, a_2, b_1 and b_2. This system is given in sys_mod01.p file.

2. A first-order system is given. This is in the form of a file sys_mod.p which is a MATLAB file.

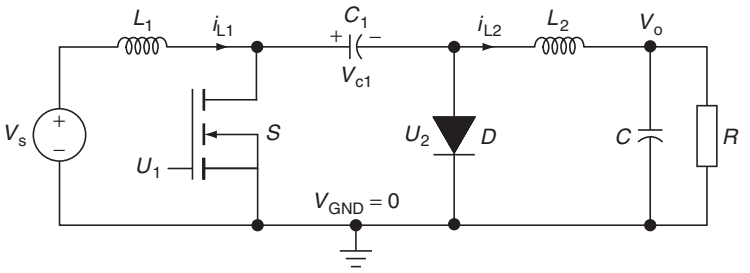


Figure 11.22 | Cuk converter circuit.

The file sys_mod.p is included in the accompanying CD. The system has the following syntax:

$$y = \text{sys\_mod}(u)$$

where u is the input and y is the output. The model of the system given in sys_mod.p is of the form

$$(1 + b_1 z^{-1}) y_k = (a_0 + a_1 z^{-1}) u_{k-1}$$

The objective is to estimate a_0, a_1 and b_1 using the least square algorithm for minimizing the error. This involves the following tasks:

a. Write a general purpose **LMS algorithm** as an *m*-function.
b. Formulate the problem for least square solution.
c. Generate input and output data for the system using random inputs.
d. Estimate the parameters.
e. Plot the evolution of error and parameters with time.

After trying out the algorithm on the above system, check if the algorithm works for another system which is of the form

$$(1 + b_1 z^{-1} + b_2 z^{-2}) y_k = (a_0 + a_1 z^{-1} + a_2 z^{-2}) u_{k-1}$$

wherein one has to estimate a_0, a_1, a_2, b_1 and b_2. This system is given in sys_mod01.p file.

3. Consider the Cuk converter DC–DC converter shown in Figure 11.22. The small signal model of the converter of Figure 11.22 is given as

$$\hat{\dot{x}} = A\hat{x} + B\hat{u}$$
$$\hat{y} = C\hat{x}$$

where

$$A = \begin{bmatrix} 0 & 0 & \dfrac{-(1-D)}{L_1} & 0 \\ 0 & 0 & \dfrac{-D}{L_2} & \dfrac{-1}{L_2} \\ \dfrac{1-D}{C_1} & \dfrac{D}{C_1} & 0 & 0 \\ 0 & \dfrac{1}{C_2} & 0 & \dfrac{-1}{RC_2} \end{bmatrix};$$

$$B = \begin{bmatrix} \dfrac{1}{L_1} & \dfrac{-V_g}{L_1(1-D)} \\ 0 & \dfrac{-V_g}{L_2(1-D)} \\ 0 & \dfrac{I_o}{C_1(1-D)} \\ 0 & 0 \end{bmatrix}; \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

The converter is designed to meet the following specification: $V_g = 130$ V, $I_0 = 1.5$ A, $V_0 = 15$ V, $f_s = 10$ kHz, $\Delta V_o = 1500$ mV and $D = 0.6$. The sampling frequency is 5 kHz. To meet the above specification, the selection of components are: $R = 10\ \Omega$, $C_1 = 100\ \mu$F, $C_2 = 10\ \mu$F, $L_1 = 10$ mH, $L_2 = 10$ mH. The plant noise co-variance matrix is

$$\begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$$

and the sensor noise co-variance is 0.1.

a. Design the Kalman estimator gain for the above system.

b. Design the linear quadratic regulator gain for the above system.

4. Consider the buck converter DC–DC converter shown in Figure 11.23. This circuit includes few non-idealities of the components of the circuit. The small signal state space model for the above circuit is given as Eq. (11.77)

$$\begin{bmatrix} \frac{d\hat{i}}{dt} \\ \frac{d\hat{v}_c}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{1}{L}\left(D(R_q + R_s) + (1-D)R_d + R_w - \frac{R_c R}{R+R_c} \right) & -\frac{1}{L}\left(\frac{R+2R_c}{R+R_c} \right) \\ \frac{R}{c(R+R_c)} & -\frac{1}{c(R+R_c)} \end{bmatrix} \begin{bmatrix} \hat{i} \\ \hat{v}_c \end{bmatrix} + \begin{bmatrix} \frac{D}{L} & \frac{V_g - V_d}{L} \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{v}_g \\ \hat{d} \end{bmatrix}$$

$C = [0\ 1]$

The following parameters are given: $V_g = 10$ V; $L = 1e{-}3$; $C = 100e{-}6$; $R = 2$; $V_d = 0.7$; $D = 0.4$; $R_c = 1e{-}3$; $R_w = 10$ mΩ; $R_q = 10$ mΩ; $R_s = 0.1$; $R_d = 0.01$; $T_s = 1/20$ kHz.

a. Design the Kalman estimator gain for the above system assuming the noise co-variances of the same order of magnitude as that for the previous exercise.

b. Design the linear quadratic regulator gain for the above system.

c. Assume 50% uncertainty in the value of C, 20% uncertainty in the value of L and 10% uncertainty in the non-ideal component values. Obtain the transfer function of the output with respect to the duty ratio. Compute the multiplicative uncertainty, $M(j\omega)$.

d. Plot the reciprocal of the maximum singular value of $M(j\omega)$ as a function of frequency. This will provide the robustness upper bound.

e. Compute the maximum singular value of the co-sensitivity $T(j\omega)$ and plot it as a function of frequency.

Note: Software packages like MATLAB give computation and plotting of singular values.

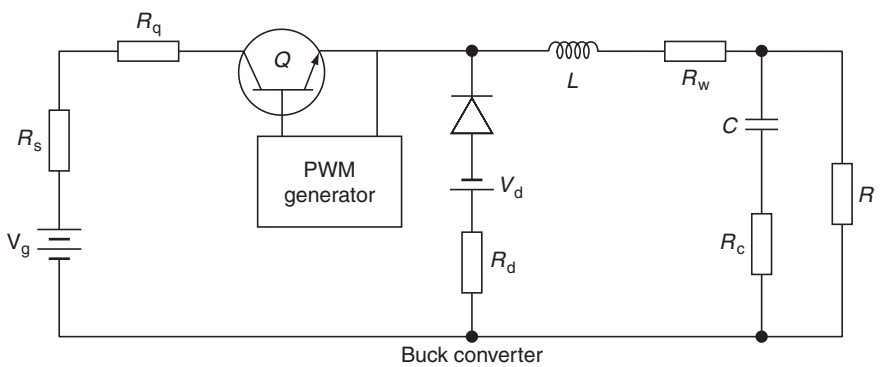


Figure 11.23 | Buck converter circuit with non-idealities.

FILL IN THE BLANKS

1. A controller that has been designed to meet the performance criteria under some constraints is called an _____ controller.
2. An estimator that has been designed to meet the performance criteria under some constraints is called an _____ estimator.
3. Kalman gain is one in an _____ controller.
4. Linear quadratic regulator achieves the control performance criteria under the constraints of minimum _____ and _____ energies.
5. Kalman filter is an _____ estimator that estimates the unavailable states of a system in the presence of _____.
6. If $\mathbf{x}$ is the state vector and e_i are errors at different instants of time, then the value of $\mathbf{x}$ that gives minimum value of $\sum_{i=1}^{n} e_i^2$ is the _____ solution.
7. Least squares solution is the _____ estimate of the state vector.
8. For all non-zero values of $\mathbf{x}$, if $\mathbf{x}^T\mathbf{C}\mathbf{x} > 0$, then $\mathbf{C}$ is called _____ definite.
9. For all non-zero values of $\mathbf{x}$, if $\mathbf{x}^T\mathbf{C}\mathbf{x} < 0$, then $\mathbf{C}$ is called _____ definite.
10. For all non-zero values of $\mathbf{x}$, if $\mathbf{x}^T\mathbf{C}\mathbf{x}$ is greater than or equal to zero, then $\mathbf{C}$ is called _____ _____-definite.
11. For all non-zero values of $\mathbf{x}$, if $\mathbf{x}^T\mathbf{C}\mathbf{x}$ is lesser than or equal to zero, then $\mathbf{C}$ is called _____ _____-definite.
12. For all non-zero values of $\mathbf{x}$, if $\mathbf{x}^T\mathbf{C}\mathbf{x}$ is either greater than, lesser than or equal to zero, then $\mathbf{C}$ is called _____.
13. For $\mathbf{x}^T\mathbf{C}\mathbf{x}$, if $\mathbf{C}$ is positive definite, then there always exists a _____.
14. For $\mathbf{x}^T\mathbf{C}\mathbf{x}$, if $\mathbf{C}$ is negative definite, then there always exists a _____.
15. All physical systems will reach a stable state also called the _____ state.
16. In the equilibrium state the system will always be at its minimum _____ state.
17. Maximum energy will be removed from a system such that it reaches the _____ or _____ energy state.
18. If the measured data has a large variance, then that data should be trusted _____.
19. If the measured data has a small variance, then that data should be trusted _____.
20. The LMS algorithm is a _____ least squares algorithm.
21. The LMS algorithm chooses the correction gain in such a way that the minimum is reached with the _____ slope.
22. The Kalman filter is a _____ estimator algorithm.
23. The Kalman filter algorithm chooses the correction gain in such a way that the error in the presence of system _____ is minimized.
24. The LMS algorithm belongs to the class of _____ estimator.
25. The Kalman filter belongs to the class of _____ estimator.
26. The steady-state Kalman gain will give _____ estimator performance.
27. In LQR, the state weighting matrix should be positive _____.
28. In LQR, the input weighting matrix should be positive _____ for a minimum to exist.
29. The actual plant and the model plant will never be _____.

30. Stability robustness is the ability to _____ closed-loop stability in spite of parameter _____.
31. Performance robustness is the ability to _____ acceptable performance even though the system may be subject to _____.
32. The multi-variable Bode magnitude plot is the plot of the transfer function _____versus the frequency.
33. For any given input, the magnitude of the transfer function at any given frequency is bounded above by its _____ value and bounded below by its _____ value.
34. The design of the robust controller needs only the maximum and minimum _____ values of the transfer function.
35. For stability robustness, the co-sensitivity should be bounded above by the _____ of the multiplicative modeling discrepancy bound.

DESCRIPTIVE QUESTIONS

1. Explain least squares principle.
2. Discuss quadratic forms. Where and why are quadratic forms used?
3. What is the minimum energy principle for any given system?
4. What is the least squares solution for $y = Cx$? Give a graphical interpretation of the least squares solution.
5. Explain the motivation for weighted least squares.
6. How are the weights of the weighting matrix in the weighted least squares method selected?
7. What is recursive least squares? Discuss the motivation for recursive least squares.
8. Write a short note on learning rate.
9. What is the cost function for the LMS algorithm?
10. What is the effect of the learning rate on the convergence speed in the LMS estimator?
11. Discuss how learning rate affects the stability of the LMS estimator.
12. What are the features of the Kalman estimator?
13. Explain the time update and measurement update in a Kalman estimator.
14. Discuss the expectation and co-variances of the plant and sensor noise.
15. What is state error co-variance? What is its importance in the Kalman estimator?
16. Distinguish between Kalman gain and steady-state Kalman gain in a Kalman filter.
17. Discuss the development of the cost function for the linear quadratic regulator.
18. Write short notes on: (a) input energy weighting and (b) state energy weighting.
19. Discuss the cost function for the discrete linear quadratic regulator for full-state feedback.
20. Discuss the cost function for the discrete linear quadratic tracker for output feedback.

21. What is linear quadratic Gaussian? Discuss.

22. Write short notes on: (a) gradient descent minimization, (b) SIMPLEX minimization.

23. What is the motivation for robust controller design?

24. Write short notes on: (a) sensitivity, (b) co-sensitivity.

25. Discuss the singular value algebra.

26. Discuss the frequency domain performance specifications for robust controller design in terms of the singular values of the transfer functions.

27. How are modeling uncertainties incorporated into the robust controller design?

28. Write short notes on: (a) additive uncertainties, (b) multiplicative uncertainties.

29. Discuss the robustness bound.

PROBLEMS

1. Consider the following deterministic system:

$$(1 + b_1 z^{-1} + b_2 z^{-2}) y_k = (a_0 + a_1 z^{-1} + a_2 z^{-2})\, u_k$$

Formulate an estimator to estimate a_0, a_1, a_2, b_1 and b_2 using the least square algorithm for minimizing the error.

2. In a particular laboratory experiment, a multimeter is used to measure the voltage at a particular node of a DC circuit. Three readings were taken at different instants of the time. The first reading was 94 V, the second 97 V and the third 89 V. What is the least squares estimate of the voltage at the node (a) after two readings, (b) after three readings?

3. Consider the following deterministic system

$$(1 + b_1 z^{-1} + b_2 z^{-2}) y_k = (a_0 + a_1 z^{-1} + a_2 z^{-2}) u_k$$

Formulate a LMS estimator to estimate a_0, a_1, a_2, b_1 and b_2.

4. What are the dimensions of the plant noise co-variance matrix and the sensor noise co-variance matrix?

5. For a first-order single-input single-output system, derive the steady-state error co-variance and the Kalman estimator gain. What is the effect of the plant and sensor noise co-variances on the Kalman estimator gain?

6. Derive the cost function for the least square regulator.

7. Suggest generic methods for obtaining the values of the state weighting matrix and the input weighting matrix for a given system.

8. For the following transfer functions compute the multiplicative uncertainty for 20% variations in the pole and zero locations:

a. $G_p(s) = \dfrac{10}{s(s+1)}$

b. $G_p(s) = \dfrac{10(s+2)}{(s+0.5)(s+1)(s+3)}$

9. For Problem 8, assume a sampling rate of 1 ms and convert the transfer functions to the discrete domain using zero-order hold. Now compute the multiplicative uncertainty for the discrete plant.

10. For Problem 9, plot the robustness bound and the co-sensitivity.

ANSWERS

Fill in the Blanks

1. optimal
2. optimal
3. optimal
4. input; state
5. optimal; noise
6. least squares
7. best
8. positive
9. negative
10. positive; semi
11. negative; semi
12. indefinite
13. minimum
14. maximum
15. equilibrium
16. energy
17. equilibrium; minimum
18. less
19. more
20. recursive
21. steepest
22. recursive
23. noise
24. prediction
25. current
26. sub-optimal
27. semi-definite
28. definite
29. identical
30. guarantee; variations
31. guarantee; disturbances
32. singular values
33. maximum singular; minimum singular
34. singular
35. reciprocal

CHAPTER 12

Discrete Computation Essentials

Learning Objectives

After reading this chapter, you will be able to:

- select the appropriate numeric format for a specific application.
- use the various arithmetic algorithms.

It must be mentioned that most of the applications today are implemented with the microcontroller or microprocessor as the controlling device. This implies that all the control algorithms will be implemented within the digital domain. This would require the implementation of equations, generation of waveforms and implementation of the controllers. This chapter discusses the systematic approach towards implementation of the numerical algorithms for the various applications.

The first step in a discrete domain implementation is the selection of an appropriate fixed-point numeric format. The number of fractional bits and the number of integer bits are selected by applying engineering logic for the specific application. One should also become familiar with the basic arithmetic like addition, subtraction and multiplication for a specific numeric format.

The next important part of discrete domain implementation is the notion of normalization and scaling. This is very important to the development of efficient and reliable working algorithms. Therefore, a proper understanding of these concepts will go a long way in developing robust implementations. Implementation of equations would involve operations like reciprocal, square root, exponential, sine and cosine functions and logarithms. The implementation of these operations will be discussed. This will be followed by a discussion on commonly implemented algorithms like PI controllers and PWM generation that are required for most power electronic systems. The discussion of the algorithms in this chapter will be by use of pseudo-codes. These pseudo-codes may be converted to the respective codes for a specific processor by appropriately mapping the instruction set to the pseudo-codes.

12.1 Numeric Formats

In the implementation of the control and signal generation algorithms for the various power electronic systems, the governing equations are based on the real numbers. The real numbers are represented as $\mathbb{R}$. The real number system $\mathbb{R}$ has the following two distinctive features:

1. $\mathbb{R}$ is infinite in range (both positive and negative). This implies that the real number line extends from 0 to $+\infty$ and 0 to $-\infty$.
2. $\mathbb{R}$ is infinitely dense. This implies that the interval between any two real numbers contains infinite number of real numbers.

However, in a digital processor, due to the limitation of the length of the data bits, the discrete real number has neither of the above two features. As a consequence, this results in two major problems:

1. **Overflow or Out-of-Range Problem:** As the data bit length is finite, the range of numbers that can be represented is also finite. Any number on the real number line that is greater than the maximum number that can be represented, will result in an overflow error.
2. **Precision or Resolution Problem:** Again as the data bit length is finite, the minimum number that can be represented is also finite. Therefore, the interval between any two real numbers contains finite number of real numbers. This implies that the resolution is finite.

Therefore, for discrete computation, one has to discuss and focus on the finite real number system that can be applied and used for implementation in a digital processor. Broadly, there are two classes of representations:

1. **Fixed-Point Number System:** It is the more popular format for low computational power and low-cost digital processors.
2. **Floating-Point Number System:** It is used in the high computational power and high-cost processors.

Fixed-Point Numeric Format

The representation of numbers in a fixed-point format is a generalization of the familiar decimal representation of a number as a string of digits with a decimal point. In this notation, the digits to the left of the decimal point represent the integer part of the number and the digits to the right of the decimal point represent the fractional part of the number. Thus, in general, a real number system is represented in fixed-point format as

$$\mathbb{D}(B, i, f)$$

where B is the base (generally 2 for binary numbers, 10 for decimal numbers); i the number of digits for the integer part; f the number of digits for the fractional part.
The real number x can be written as

$$x = \pm\left[\left(d_0 \cdot B^0 + d_1 \cdot B^1 + \cdots + d_{i-1} \cdot B^{i-1}\right) + \left(\frac{d_{-1}}{B} + \frac{d_{-2}}{B^2} + \cdots + \frac{d_{-f}}{B^f}\right)\right]$$

EXAMPLE 12.1

For a 16-bit processor

$\mathbb{D}(2, 0, 16)$ implies base 2 system with 0 sign bit, 0 bit for integer digit, 16 bits for the fractional digits

$\mathbb{D}(2, 0, 15)$ implies base 2 system with 1 sign bit, 0 bit for integer digit, 15 bits for fractional digits.

$\mathbb{D}(2, 4, 12)$ implies base 2 system with no sign bit, 4 bits for integer digits, 12 bits for fractional digits.

$\mathbb{D}(2, 3, 12)$ implies base 2 system with 1 sign bit, 3 bits for integer digits, 12 bits for fractional digits.

which is written as

$$x = \pm d_{i-1} \ldots d_1 d_0 \, . \, d_{-1} d_{-2} d_{-3} \ldots d_{-f}$$

$$\text{where } 0 \le d_k < B \text{ for } -f \le k < i.$$

Relative Error and Relative Density

The relative error is defined as

$$r_E = \left(\frac{x - \text{fix}(x)}{x} \right) \times 100 \tag{12.1}$$

where r_E is the percent relative error; x the real number; fix(x) the real number that is represented in the specified digital domain fixed-point format that is fix(x) $\in \mathbb{D}(B, i, f)$.

Consider a five-digit fixed-point system $\mathbb{D}(10, 3, 1)$. This is a base 10 system that consists of 1 sign digit, 3 integer digits and 1 fractional digit. With this system the range of numbers that can be represented is from –999.9 to +999.9. A real number x in this range can be represented by

$$\text{fix}(x) \in \mathbb{D}(10, 3, 1)$$

Consider a representative real number towards the end of the range, say $x = 865.54$. Then $\text{fix}(x) = 865.5$. The absolute error is $|x - \text{fix}(x)|$ and is equal to 0.04. From Eq. (12.1), relative error is given as

$$r_E = \left(\frac{865.54 - 865.5}{865.54} \right) \times 100 \approx 0.005\%$$

If a representative real number near zero is selected, say $x = 0.86554$, then fix(x) = 0.9. The absolute error is less than 0.05. The relative error is

$$r_E = \left(\frac{0.86554 - 0.9}{0.86554} \right) \times 100 \approx -4\%$$

This shows that the relative error near zero is more than the relative error near the end of the range for similar absolute errors. This implies that the *relative density* of the fixed-point number system is not uniform within the range. This, in fact, is the basic disadvantage of the fixed-point number system.

Problem 12.1

Consider an eight-bit digital processor. If one bit is used as sign bit, then what is the range of numbers for

(a) $\mathbb{D}(2, 3, 4)$

(b) $\mathbb{D}(2, 1, 6)$

Solution

(a) The fixed-point system $\mathbb{D}(2, 3, 4)$ is a binary system (i.e., base 2) with the fractional part having four bits. Out of the remaining four bits, one bit is used as the sign bit. The remaining three bits are the integer bits. A number x is represented as

$$\underset{\text{sign}(2^3)}{\square}\quad \underbrace{\square\ \square\ \square}_{\text{Integer part}}\ \cdot\ \underbrace{\square\ \square\ \square\ \square}_{\text{Fractional part}}$$
$$\qquad\quad 2^2\ \ 2^1\ \ 2^0 \qquad\quad 2^{-1}\ \ 2^{-2}\ \ 2^{-3}\ \ 2^{-4}$$

The largest number is 0 111 1111. This is the representation of

$$x = \left(-\text{sign}\cdot 2^3\right) + \left(2^2 + 2^1 + 2^0\right) + \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}\right)$$
$$= \quad 0 \quad + \quad 7 \quad + \quad 0.9375 \quad = 7.9375$$

Thus, $(01111111)_2 \Leftrightarrow (7F)_{16} \Leftrightarrow (7.9375)_{10}$. This binary base 2 number can be equivalently represented in base 16 (or hexadecimal) as $7F$ or in base 10 (decimal) as 7.9375.

For the negative number, the 2's complement is used. The 2's complement of the largest number gives the minimum value that the above system can represent. Thus, taking the 2's complement of the above number, one obtains

2's complement of 0 111 1111 = 1 000 0001

which is the representation of the following base 10 number:

$$x = (-\text{sign}\cdot 2^3) + (0\cdot 2^2 + 0\cdot 2^1 + 0\cdot 2^0) + \left(\frac{0}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \frac{1}{2^4}\right)$$
$$= \quad -8 \quad + \quad 0 \quad + \quad 0.0625 \quad = -7.9375$$

(b) The fixed-point system $\mathbb{D}(2, 1, 6)$ is a binary system with the fractional part having six bits. Out of the remaining two bits, one bit is used as sign bit. The remaining one bit is the integer bit. A number x is represented as

$$\underset{\text{sign}(2^1)}{\square}\quad \underbrace{\square}_{\text{Integer part}}\ \cdot\ \underbrace{\square\ \square\ \square\ \square\ \square\ \square}_{\text{Fractional part}}$$
$$\qquad\quad 2^0 \qquad\quad 2^{-1}\ \ 2^{-2}\ \ 2^{-3}\ \ 2^{-4}\ \ 2^{-5}\ \ 2^{-6}$$

The largest number is 0 1 111111. This is the representation of the following base 10 number:

$$x = (-\text{sign}\cdot 2^1) + (2^0) + \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^6}\right)$$
$$= \quad 0 \quad + \quad 1 \quad + \quad 0.984375 \quad = 1.984375$$

The 2's complement of the largest number gives the minimum value that the above system can represent. Thus, taking the 2's complement of the above number, one obtains

2's complement of 0 1 111111 = 1 0 000001

which is the representation of the following base 10 number:

$$x = (-\text{sign}\cdot 2^1) + (0\cdot 2^0) + \left(\frac{0}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \frac{0}{2^4} + \frac{0}{2^5} + \frac{1}{2^6}\right)$$
$$= \quad -2 \quad + \quad 0 \quad + \quad 0.015625 \quad = -1.934375$$

Thus, $(10000001)_2 \Leftrightarrow (81)_{16} \Leftrightarrow (-1.934375)_{10}$.

Floating-Point Numeric Format

The floating-point representation consists of a mantissa M which is multiplied by the exponential factor B^E, where B is the base and E is the exponent. A real number system in floating-point format is represented as

$$\mathbb{D}(B, f, L, U)$$

where B is the number base; f the precision of the mantissa or length of the fractional part of the mantissa; L the lower bound of the exponent E and U the upper bound of the exponent E.

A non-zero real number x is represented in the floating-point notation as

$$x = M \cdot B^E \tag{12.2}$$

where $L \le E \le U$.

The real number x can be written as

$$x = \pm\left(\frac{d_1}{B} + \frac{d_2}{B^2} + \frac{d_3}{B^3} + \cdots + \frac{d_f}{B^f}\right) \times B^E$$

which is written as

$$x = \pm 0.d_1 d_2 d_3 \ldots d_f \times B^E$$

where $1 \le d_1 < B$ (implies that the first digit is non-zero); $0 \le d_k < B$ for $2 \le k \le f$.

The above representation where d_1 is non-zero is called the *normalized form*. In the normalized form, the absolute value of mantissa (M) is

$$\frac{1}{B} \le M < 1$$

However, in the normalized form, one cannot represent the number zero directly. It can be represented as the subtraction of two equal normalized numbers. Normalized representation of the numbers in the floating-point system improves the precision as can be observed from the following example.

EXAMPLE 12.2

Consider a floating-point system $\mathbb{D}(10, 4, -2, 3)$. With this system the range of numbers that can be represented in the normalized form is from $\pm 0.1 \times 10^{-2}$ to $\pm 0.9999 \times 10^{3}$ (i.e., -999.9 to -0.001 and $+0.001$ to $+999.9$).
Let $x = 0.029411765$ which can be represented in the following two forms:

1. **Without normalization:** 0.0294×10^{0}
2. **With normalization:** 0.2941×10^{-1}

It can be observed that the precision in the normalized representation is more than that without normalization. The digits are better utilized with normalization.

Relative Error and Relative Density

The relative error is defined as

$$r_E = \left(\frac{x - \text{float}(x)}{x}\right) \times 100 \tag{12.3}$$

where r_E is the percent relative error, x the real number, float(x) is the real number that is represented in the specified digital domain floating-point format, that is, float(x) $\in \mathbb{D}(B, f, L, U)$.

Consider a floating-point system $\mathbb{D}(10, 4, -2, 3)$. Let $x = 865.54$. Then $\text{float}(x) = 0.8655 \times 10^3$. From Eq. (12.3), the relative error is given as

$$r_E = \left(\frac{865.54 - 865.5}{865.54} \right) \times 100 \approx 0.005\%$$

If a representative real number near zero is selected, say $x = 0.86554$, then $\text{float}(x) = 0.8655 \times 10^0$. The relative error is

$$r_E = \left(\frac{0.86554 - 0.8655}{0.86554} \right) \times 100 \approx 0.005\%$$

This shows that the relative error near zero is the same as the relative error near the end of the range for similar absolute errors. This implies that the relative density of the floating-point number system is uniform within the range. Thus, the floating-point system has *uniform relative density*.

However, in the case of normalized floating-point system, there exists an underflow hole, that is, there is a jump from minimum value to zero, which implies that values between the minimum value and zero cannot be represented. The underflow hole is depicted in Figure 12.1(b). Figure 12.1(a) shows the pictorial depiction of the fixed-point system and Figure 12.1(b) shows the pictorial depiction of the floating-point system. In the fixed-point system the numbers are uniformly distributed over the range. This makes the relative density non-uniform in the fixed-point system. However in the floating-point system, the numbers have a higher density towards zero as compared to the numbers away from zero. This makes the relative density uniform in the floating-point system.

In Figure 12.1(b), the underflow hole can be observed. To take care of the underflow hole problem, the floating-point normalized representation is generally modified as

$$x = \pm 1.d_1 d_2 d_3 \ldots d_f \times B^E$$

where $0 \le d_k < B$ for $1 \le k \le f$ and $0 = 1.00 \times B^0 - 1.00 \times B^0$. In this representation, there is no restriction on d_1 to be non-zero. Therefore one can represent smaller numbers. For example, in the floating-point

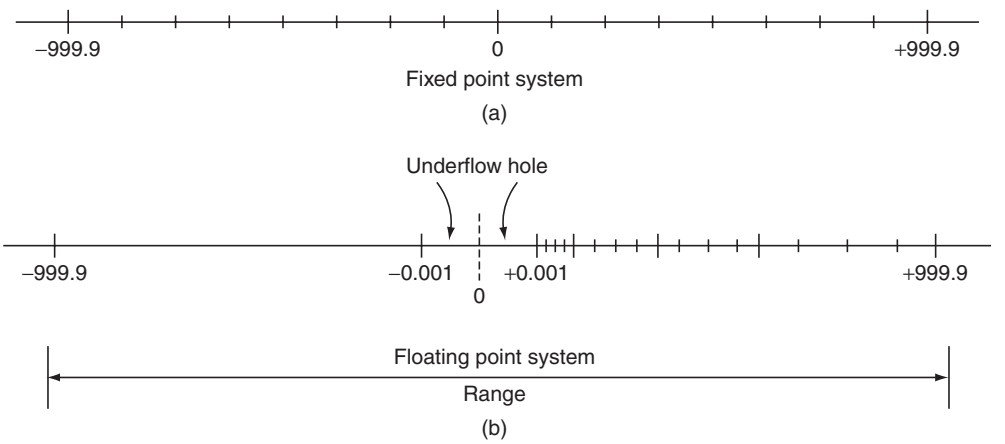


Figure 12.1 | Pictorial depiction of: (a) Fixed-point system; (b) floating-point system.

system $\mathbb{D}(10, 4, -2, 3)$ the minimum in the previous normalized representation is 0.001. If $x = 0.0001$, then it cannot be represented. In the modified normalized representation, $x = 0.0001$ is obtained as

$$\text{float}(x) = 1.0002 \times B^0 - 1.0001 \times B^0$$

Problem 12.2

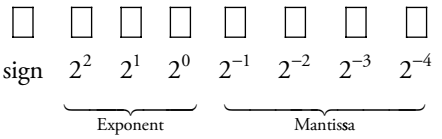
Consider an eight-bit digital processor. If one bit is used as sign bit, then what is the range of numbers for

(a) $\mathbb{D}(2, 4, 0, 7)$

(b) $\mathbb{D}(2, 4, -3, 3)$

Solution

(a) The floating-point system $\mathbb{D}(2, 4, 0, 7)$ is a binary system (i.e., base 2) with the mantissa having four bits. Out of the remaining four bits, one bit is used as sign bit. The remaining three bits are the exponent bits. It should be noted that the sign bit is for the mantissa part alone. A number x is represented as



The smallest positive number is **0** 000 **1000**. This is the representation of the following base 10 number,

$$x = 0.5 \times 2^0 = 0.5$$

The largest positive number is **0** 111 **1111**. This is the representation of the following base 10 number:

$$x = \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}\right) \times 2^7 = 0.9375 \times 2^7 = 120$$

For negative numbers, the 2's complement of the mantissa alone is performed. For the number **0** 000 **1000**, on taking the 2's complement of the mantissa along with the sign bit, the bit pattern **1** 000 **0001** is obtained and this represents

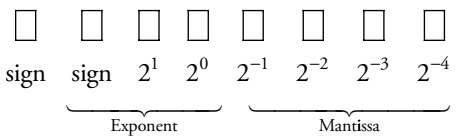
$$x = \left[(-\text{sign} \cdot 2^0) + \left(\frac{0}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \frac{1}{2^4}\right)\right] \times 2^7 = [-1 + 0.0625] \times 2^7 = -0.9375 \times 2^7 = -120$$

However, it should be noticed that the mantissa in the above case is not normalized. In the normalized form, the first digit of the mantissa should always be non-zero. Thus, the maximum negative number that can be represented in the normalized form is **1** 000 **1001**. This represents the following base 10 number,

$$x = \left[(-\text{sign} \cdot 2^0) + \left(\frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \frac{1}{2^4}\right)\right] \times 2^7 = [-1 + 0.5625] \times 2^7 = -0.4375 \times 2^7 = -56$$

As the mantissa is normalized, the sign bit value will be the next higher bit value with respect to the mantissa and therefore the bit value of the sign bit is 2^0 as indicated in the above calculation.

(b) The floating-point system $\mathbb{D}(2, 4, -3, 3)$ is a binary system (i.e., base 2) with the mantissa having four bits. Out of the remaining four bits, one bit is used as sign bit. The remaining three bits are the exponent bits. Among the three exponent bits, one is used as sign bit for the exponent. This will leave only two bits for the exponent amplitude. A number x is represented as



The smallest positive number is **0** 101 **1000**. Here the exponent part alone is the 2's complement of 011 which is 101 or −3. This is the representation of $x = 0.5 \times 2^{-3} = 0.0625$.
The largest positive number is **0** 011 **1111**. This is the representation of the following base 10 number:

$$x = \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}\right) \times 2^3 = 0.9375 \times 2^3 = 7.5$$

The bit pattern 1 011 1001 represents $x = -3.5$.

General Notes on Number System

1. The relative density is uniform in the floating-point number system whereas it is not uniform in the fixed-point number system.
2. In the fixed-point system, one must keep track of the position of the point separating the integer and fractional parts whereas there is no need to keep track of the decimal or binary point in the floating-point system.
3. There is a limitation in the floating-point system in representing zero in the normalized format whereas in the fixed-point system, the representation of zero is like natural numbers.
4. The floating-point system has the problem of underflow "hole" that limits the representation of very small numbers.
5. In floating-point system, it can be observed from Figure 12.1(b) that the interval between numbers increases as the numbers grow in magnitude. It is this characteristic that allows floating-point numbers to preserve the significant digits. However, the quantizing errors are proportional to the magnitude of the number, x being represented. For normalized floating-point numbers, the worst case quantization error bounds can be expressed as

$$\frac{|\Delta x|}{|x|} \leq \in \quad \text{if truncation is employed}$$

$$\frac{|\Delta x|}{|x|} \leq \frac{\in}{2} \quad \text{if rounding is employed}$$

where $\in = B^{1-f}$ is called the machine epsilon and Δx is the error due to quantization.
6. The computation overhead for the fixed-point system is much lower than that for the floating-point number system as arithmetic operations must be performed on both mantissa and exponent. Therefore, for a given calculation, the fixed-point system is faster on a given digital processor. This would imply that for a given algorithm, a lower power digital processor can be used in a fixed-point system as compared to the floating-point system. This would reduce the cost of the algorithm and the system cost. Therefore the fixed-point system is the more popular choice. Floating-point system is used only when the computations become very large and complex.

12.2 Tracking the Base Point in the Fixed-Point System

The discussions in the remaining portion of the chapter will use the fixed-point number system as this is the first choice for a given application. Only if it becomes extremely difficult or impossible to manage with fixed-point number system should one shift to the floating-point number format. One of the major problems that is faced in the fixed-point system is the need to keep track of the virtual base point (decimal point in base 10 system or the binary point in the base 2 system). This section will discuss the primary arithmetic operations and the method of tracking the base point in a systematic manner.

There are three primary operations based on which all arithmetic calculations are performed in the digital processor. The three primary operations are:

1. addition with symbolic operator "+";
2. subtraction with symbolic operator "–";
3. multiplication with symbolic operator "*".

Addition of Numbers

The addition can be performed with two numbers only if the two numbers are in the same fixed-point numeric format. Thus, if x_1 is in $\mathbb{D}(B, i, f)$ and x_2 is in $\mathbb{D}(B, i, f)$, then $x_1 + x_2$ is also in $\mathbb{D}(B, i, f)$. If x_1 is in $\mathbb{D}(B, i_1, f_1)$ and x_2 is in $\mathbb{D}(B, i_2, f_2)$ where $i_1 \neq i_2$ and $f_1 \neq f_2$ then it is not possible to add x_1 and x_2. Either x_1 should be converted to $\mathbb{D}(B, i_2, f_2)$ or x_2 should be converted to $\mathbb{D}(B, i_1, f_1)$ format before performing the addition. The conversion from one numeric format to another for the same base B is done by either right shifting or left shifting.

Subtraction of Numbers

Like addition, subtraction also can be performed with two numbers only if the two numbers are in the same fixed-point numeric format. Thus, if x_1 is in $\mathbb{D}(B, i, f)$ and x_2 is in $\mathbb{D}(B, i, f)$, then $x_1 - x_2$ is also in $\mathbb{D}(B, i, f)$. If x_1 is in $\mathbb{D}(B, i_1, f_1)$ and x_2 is in $\mathbb{D}(B, i_2, f_2)$ where $i_1 \neq i_2$ and $f_1 \neq f_2$ then it is not possible to subtract x_2 from x_1 without converting one of the numbers into a compatible numeric format.

Multiplication of Numbers

Unlike addition or subtraction, multiplication can be performed on two numbers of different numeric formats but with the same base B. Thus if x_1 is in $\mathbb{D}(B, i_1, f_1)$ and x_2 is in $\mathbb{D}(B, i_2, f_2)$, then $x_1 \times x_2$ will be in $\mathbb{D}(B, i_1 + i_2, f_1 + f_2)$.

EXAMPLE 12.3

Let $x_1 = +113.54$ in $\mathbb{D}(10, 3, 2)$ and $x_2 = +008.34$ in $\mathbb{D}(10, 3, 2)$. The product is given as

$$x_1 \times x_2 = (+113.54) \times (+008.34) = +000946.9236$$

Observe that the product is in $\mathbb{D}(10, 6, 4)$. If the product has to be converted to $\mathbb{D}(10, 3, 2)$ numeric format, then there are two possible methods:

1. **Shift Right:** The product $x_1 \times x_2$ in $\mathbb{D}(10, 6, 4)$ is shifted right by two digits. This will cause the two lower significant digits to be removed and two zeros will be included in the most significant bits (MSBs). The number now is +00000946.92. The 5 least significant bits (LSBs) of this number are stored in a new variable as +946.92 which is in $\mathbb{D}(10, 3, 2)$.

2. **Shift Left:** The product $x_1 \times x_2$ in $\mathbb{D}(10, 6, 4)$ is shifted left by three digits. This will result in +946.9236000. The 5 MSBs of the left-shifted number are stored in a new variable as +946.92 which is in $\mathbb{D}(10, 3, 2)$.

In general if x_1 is in $\mathbb{D}(B, i_1, f_1)$ and x_2 is in $\mathbb{D}(B, i_2, f_2)$, then let the product $x_1 \times x_2$ be in $\mathbb{D}(B, i_p, f_p)$. To convert the product to $\mathbb{D}(B, i_1, f_1)$, then shift right by f_2 number of bits and extract the $(i_1 + f_1)$ LSBs or shift left by i_2 number of bits and extract the $(i_1 + f_1)$ MSBs to express the product in $\mathbb{D}(B, i_1, f_1)$. In order to express the product in $\mathbb{D}(B, i_2, f_2)$, shift right by f_1 number of bits and extract the $(i_2 + f_2)$ LSBs or shift left by i_1 number of bits and extract the $(i_2 + f_2)$ MSBs.

EXAMPLE 12.4

Let $x_1 = +1213.54$ in $\mathbb{D}(10, 4, 2)$ and $x_2 = +04.347$ in $\mathbb{D}(10, 2, 3)$. The product is given as

$$x_1 \times x_2 = (+1213.54) \times (+04.347) = +005275.25838$$

The product $x_1 \times x_2$ is in $\mathbb{D}(10, 6, 5)$. If the product $x_1 \times x_2$ is to be represented in $\mathbb{D}(10, 4, 2)$ say, then $i_1 = 4$, $f_1 = 2$, $i_2 = 2$ and $f_2 = 3$. Shift right the product number by f_2 bits (three bits), that is, +000005275.25 and extract the $(i_1 + f_1)$ $(4 + 2 = 6)$ LSBs, that is, +5275.25. Alternately left shift the product number by i_2 bits (two bits), that is, +5275.2583800 and extract the $i_1 + f_1$ $(4 + 2 = 6)$ MSBs, that is, +5275.25.

12.3 Normalization and Scaling

For an n-digit system, the numbers are represented in a numeric format $\mathbb{D}(B, i, f)$. The resolution of the numeric format is $1/B^f$. The precision of the numbers is highest if "f" has the maximum number of digits allotted to it out of the available n digits. In an n-digit system, f is maximum when

$$i = 0 \text{ and } f = n, \text{ that is, } \mathbb{D}(B, 0, n)$$

For highest precision, the resolution is $1/B^n$. However, in the digital domain with finite number of digits, one has to compromise between the range and the precision. If the precision is to be high, then the number of integer bits will become low and as a consequence the range will reduce. In the limiting case when the fractional part $f = n$, the integer part $i = 0$. Thus for an n-digit system, at maximum precision, the range is limited to 1.

If $\mathbb{D}(B, 0, n)$ numeric format is to be used, then all the variables in the processor for that specific application should have a value less than 1 so that the values are within the range 0–1. It is not sufficient if the values of the variables are within the range. The values resulting from the primary arithmetic operations must also stay within the range. Thus if x_1 and x_2 are in $\mathbb{D}(B, 0, n)$, then

1. $x_1 + x_2 < 1$ if $x_{1\max} = a$ then $x_{2\max} < 1 - a$ and vice-versa where $a < 1$.
2. $x_1 - x_2 < 1$ if $x_1 > x_2$. If $x_2 > x_1$, then the result will be out of range as there is no sign bit in $\mathbb{D}(B, 0, n)$ numeric format.
3. $x_1 \times x_2 < 1$ if both x_1 and x_2 are less than 1.

Therefore, $\mathbb{D}(B, 0, n)$ is not a very practically useful numeric format. A popular numeric format that can cater to many applications is $\mathbb{D}(B, 1, n - 2)$ for an n-digit system. This implies that there is 1 sign digit, 1 integer digit

and $n-2$ fractional digits giving a precision of $1/B^{n-2}$. For a binary system, B is 2 and this numeric format has a range from –1.9999… to +1.9999… (approximately, from –2 to +2). If x_1 and x_2 are in $\mathbb{D}(2, 1, n-2)$, then

1. $x_1 + x_2$ should be < 2 if $x_1 < 1$ and $x_2 < 1$.
2. $x_1 - x_2$ should be $< \pm 2$ if $x_1 < 1$ and $x_2 < 1$.
3. $x_1 \times x_2 < 1$ if both x_1 and x_2 are less than 1.

EXAMPLE 12.5

Consider a 16-bit system with $\mathbb{D}(2, 1, 14)$ numeric format. This implies a sign bit also. The characteristic features of this 16-bit numeric format are

1. **Range:** $(1000\ 0000\ 0000\ 0001)_2$ to $(0111\ 1111\ 1111\ 1111)_2$ or $(8001)_{16}$ to $(7FFF)_{16}$ or $(-1.9999)_{10}$ to $(+1.9999)_{10}$.
2. If x_1 and x_2 are in $\mathbb{D}(2, 1, 14)$, then x_1 and $x_2 < (0011\ 1111\ 1111\ 1111)_2$ or $(3FFF)_{16}$ or $(0.9999)_{10}$ so that $x_1 + x_2 < 2$, $x_1 - x_2 < \pm 2$ and $x_1 \times x_2 < 1$ are satisfied.

Thus, from the above example, it is evident that if the variables in the application are ensured to have a value less than unity under all conditions, then all values including the result will be in the range of $\mathbb{D}(2, 1, 14)$. To achieve this, the variables are divided by the worst case maximum value that the variables can assume in a given applications. This will ensure that the variables will all be per unit values or normalized values. This is called *normalization*. For the normalized system, 2^f will represent 1 or 100%. In the above example, 2^{14} will represent 1.

Let x be a variable used within the digital processor. If x_{max} is the worst case maximum value that the variable can assume, then

$$x_n = \frac{x}{x_{max}} \tag{12.4}$$

where x_n is the normalized value of x. Consider a situation wherein an analog variable is to be discretized and taken into the digital domain. The block schematic is shown in Figure 12.2. Let the range of analog input to the analog-to-digital converter (ADC) be $0–V_{AD}$. The maximum possible input voltage to the ADC is V_{AD}. Define a base value, $x_{base} = 1$ per unit (pu) or 100%. The value of x_{base} is in the range $0 < x_{base} \leq V_{AD}$.

The value $x_{base} < V_{AD}$ is used in applications wherein some portion of the signal amplitude space is needed for representing the overloads. This is achieved by sacrificing precision under normal operating conditions. If $x_{base} = V_{AD}$, then it implies that there is no overload. The maximum value of the signal that can be represented is the rated value or the base value itself.

The output of the ADC is a binary word of n_a-bits length. For DC signals, one can use all the n_a-bits to represent the magnitude of the analog signal x. After ADC, the digital value is called $x_{AD}(k)$ at time instant kT where T is the sample time. The output of the ADC is scaled by K to obtain the n-bit normalized data $x_n(k)$ in $\mathbb{D}(2, i, f)$ numeric format.

The relationship between the input and the output of the ADC is as follows:

1. 2^{n_a} in digital domain at the output of the ADC represents V_{AD} value in the analog domain at the input of the ADC.
2. $x_{AD}(k)$ in the digital domain at the output of the ADC represents a value $x_{AD}(k) \cdot (V_{AD} / 2^{n_a})$ which is the quantized value of the analog variable x.

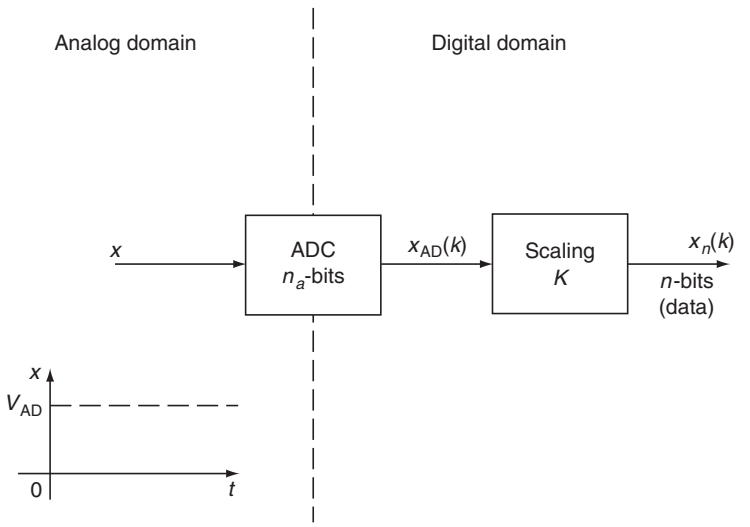


Figure 12.2 | Scaling in analog-to-digital domain conversion of variables.

The ADC output bits n_a is usually lesser than the number of data bits n. Therefore, the output of the ADC is padded with $(n - n_a)$ zero MSBs so that the ADC output is converted to n-bit data.

If x_{base} is the per unit value or rated value or the nominal value of the variable x, then to normalize, divide the variable with the base value and multiply with 2^f to convert to $\mathbb{D}(2, i, f)$ numeric format:

$$x_n = \frac{x}{x_{base}}$$

$$x_n(k) = x_n \cdot 2^f = \frac{x_{AD}(k) \cdot (V_{AD} / 2^{n_a})}{x_{base}} \cdot 2^f \tag{12.5}$$

From Eq. (12.5) it can be observed that when $x = x_{base}$, then $x_n = 2^f$ which represent 1 or 100%. Equation (12.5) can be re-written as

$$x_n(k) = x_{AD}(k)\left(\frac{V_{AD}}{x_{base}} \cdot \frac{2^f}{2^{n_a}}\right) = x_{AD}(k) \cdot K \tag{12.6}$$

The factor K in Eq. (12.6) is called the *scaling factor*.

For applications where x is an AC waveform, then x will swing both positive and negative. However, the ADC in most cases will accept only unidirectional or DC voltages. Therefore, the AC waveform, x is level shifted by $V_{AD}/2$ such that the peak-to-peak amplitude of x is within V_{AD}. In such cases, the input to the ADC is $x + (V_{AD} / 2)$. The ADC output $x_{AD}(k)$ will also be level shifted by $2^{n_a - 1}$ which represents $V_{AD}/2$. In this case, one bit of the ADC, that is, the MSB is used as a sign bit. $2^{n_a - 1}$ is subtracted from $x_{AD}(k)$ to represent x. However, it should be noted that as one of the ADC is used as sign bit, only $n_a - 1$ bits are available for representing the amplitude of x. From Eq. (12.6), the normalized quantized value of x is

$$
\begin{aligned}
x_n(k) &= [x_{\text{AD}}(k) - 2^{n_a-1}] \left(\frac{V_{\text{AD}}/2}{x_{\text{base}}} \cdot \frac{2^f}{2^{n_a-1}} \right) \\
&= [x_{\text{AD}}(k) - 2^{n_a-1}] \left(\frac{V_{\text{AD}}}{x_{\text{base}}} \cdot \frac{2^f}{2^{n_a}} \right) \\
&= [x_{\text{AD}}(k) - 2^{n_a-1}]\, K
\end{aligned}
\tag{12.7}
$$

Problem 12.3

Consider a 10-bit ADC that has an input range of 5 V. For 16-bit data to be represented in $\mathbb{D}(2, 3, 12)$ numeric format, find the scaling factor K for

(a) DC input where $x_{base} = 3$ V
(b) AC input where $x_{base} = 2.5$ V is the peak value of the AC waveform

Solution

(a) For a DC input, $V_{\text{AD}} = 5$ V, $n_a = 10$, $f = 12$, $n = 16$ and $x_{\text{base}} = 3$ V. From Eq. (12.6) one obtains

$$
x_n(k) = x_{\text{AD}}(k) \left(\frac{V_{\text{AD}}}{x_{\text{base}}} \cdot \frac{2^f}{2^{n_a}} \right) = x_{\text{AD}}(k) \left(\frac{5}{3} \cdot \frac{2^{12}}{2^{10}} \right) = x_{\text{AD}}(k) \cdot 6.67
$$

$K = (6.67)_{10}$

Atleast three bits are required to represent the integer part of K; therefore K can be expressed in $\mathbb{D}(2, 3, 12)$ numeric format.

(b) For an AC input, $V_{\text{AD}} = 5$ V, $n_a = 10$, $f = 12$, $n = 16$ and $x_{\text{base}} = 2.5$ V. From Eq. (12.7) one obtains

$$
x_n(k) = (x_{\text{AD}}(k) - 2^9) \left(\frac{5}{2.5} \cdot \frac{2^{12}}{2^{10}} \right) = (x_{\text{AD}}(k) - 2^9) \cdot 8
$$

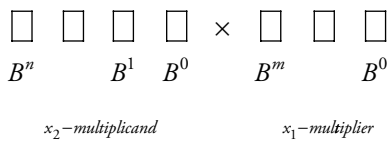
$K = (8)_{10}$

Atleast four bits are required to represent the integer part of K; therefore K can be expressed in $\mathbb{D}(2, 4, 11)$ or $\mathbb{D}(2, 5\ 10)$ numeric format.

For power electronic applications, generally, the voltage and current overshoots up to twice rated are possible in most cases for short times. This implies that the instantaneous powers are up to four times rated. This implies that at least three bits are required to represent power up to four times rated. Thus, three integer bits are normally used which can represent upto seven times the nominal value. One bit is needed for sign bit. Therefore, for a 16-bit digital system, $\mathbb{D}(2, 3, 12)$ numeric format is most appropriate for the power electronic applications.

Multiplication Algorithm

In some microcontrollers, only addition and subtraction are provided. In that case, a routine for multiplication has to be used. The multiplication uses two variables as inputs called the multiplicand and the multiplier. Let x_1 in $\mathbb{D}(B, i_1, f_1)$ be the multiplier with m digits and x_2 in $\mathbb{D}(B, i_2, f_2)$ be the multiplicand with n digits. Then the multiplier operation is shown in the following template.



The digits of x_1 are multiplied with the digits of x_2 one by one. If the digit number i of x_1 is multiplied with digit number j of x_2, then the *digit product* is added to the previous product after multiplying by B^{i+j} or added to the previous product after shifting left by $i + j$ digits. Thus,

$$x_p = x_1 \times x_2 = x_{prev} + (\text{digit product}) \times B^{i+j}$$

The above equation has to be applied till all the digits of the multiplier have been multiplied with all the digits of the multiplicand. Thus, the generic algorithm for multiplication is

```
set x_prev = 0
Loop for i = 0 to m
    Loop for j = 0 to n
    x_p = x_prev + (d_j × d_i)*B^(i+j)
    end loop j
end loop i
```

Problem 12.4

Multiply $(34)_{10}$ and $(23)_{10}$

Solution

Step 1: $x_{prev} = 0$; $x_p = 0 + (4 \times 3)*10^{0+0} = 12$
Step 2: $x_{prev} = 12$; $x_p = 12 + (3 \times 3)*10^{0+1} = 102$
Step 3: $x_{prev} = 102$; $x_p = 102 + (4 \times 2)*10^{1+0} = 182$
Step 4: $x_{prev} = 222$; $x_p = 182 + (3 \times 2)*10^{1+1} = 782$

If the base $B = 2$, then multiplication consists of only shift left and add. The above algorithm for binary multiplication becomes

```
set x_prev = 0
Loop for i = 0 to m
    Loop for j = 0 to n
        x_p = x_prev + (d_j × d_i) * 2^(i+j)
    end loop j
end loop i
```

2^{i+j} is nothing but shift left operation. The *digit product* can be either 0 or 1 in the case of the binary system. Therefore, only if the *digit product* is 1 then, shift left by $(i + j)$ and add to the x_{prev}.

12.4 Arithmetic Algorithms

Most digital processors have only three primary arithmetic operators: addition, subtraction and multiplication. If multiplication instruction is not available, then a multiplication routine has to be used as discussed in the previous section. Some advanced processors also have hardwired division that is in-built. This section discusses the methods of computing the functions in an iterative manner using only these three primary arithmetic operators.

Given a continuous function $y = f(x)$, it is required to compute the value of the function for a given value of the parameter x. If the function $y = f(x)$ is difficult to compute, it can be re-written in implicit form as $F(x, y) = 0$.

Let y_k be the approximate value of the function. If the error in the function estimate is Δy then

$$F(x, y_k) = \Delta y \cdot \left.\frac{\partial F}{\partial y}\right|_{x, y_k}$$

$$= (y_k - y_{k+1}) \cdot \left.\frac{\partial F}{\partial y}\right|_{x, y_k} \tag{12.8}$$

and the convergence is ensured if $\partial F/\partial y$ and $\partial^2 F/\partial y^2$ have same signs in the interval under consideration wherein the root of y lies. Generally, in iterative algorithms, the initial value y_0 is arbitrarily chosen such that it is as close as possible to the final value of y. The process of iteration is continued till y_k and y_{k-1} coincide or differ by a very small value $\in$.

From Eq. (12.8), the general iterative equation for a continuous function $y = f(x)$ is

$$y_{k+1} = y_k - \frac{F}{(\partial F / \partial y)\big|_{x, y_k}} \tag{12.9}$$

The advantage of the iterative algorithm is that the operations are of the same type for each iteration and therefore are easy to program. It should, however, be noted that $F(x, y) = 0$ for the function $y = f(x)$ may be realized in many ways. This fact should be utilized to obtain an implicit function that leads to rapid convergence.

Reciprocal

Let $y = 1/x$ where $x > 0$. Set

$$F(x, y) \equiv x - \frac{1}{y} = 0$$

$$\frac{\partial F}{\partial y} = F' = \frac{1}{y^2}$$

Applying Eq. (12.9), one obtains

$$y_{k+1} = y_k - \left(\frac{x - (1/y_k)}{1/y_k^2} \right)$$

Re-arranging the above, the iterative solution for the reciprocal of x is given as

$$y_{k+1} = y_k(2 - x \cdot y_k) \text{ where } k = 0, 1, 2, \ldots \tag{12.10}$$

Equation (12.10) is the iterative equation for computing the reciprocal of x. The initial value y_0 is selected as follows.

1. Let x be an n-digit binary number in $\mathbb{D}(2, i, f)$ numeric format. The binary number x can be written as

$$x = 2^m \cdot x_m \text{ where } m \text{ is an integer and } (1/2) \le x_m < 1$$

2. Check the bit number of the MSB of x. Let this be n_{msb}. Then

$$m = n_{\text{msb}} + 1$$

3. Set $y_0 = 2^{-m}$ in order to have rapid convergence.

EXAMPLE 12.6

Consider a 16-bit binary number $x = (3)_{10}$ in $\mathbb{D}(2, 3, 12)$ numeric format:

$$x = 0011.000000000000$$

(Note that the binary point is a virtual point.)
Now $n_{\text{msb}} = 1$ (corresponding to 2^1). Therefore

$$m = n_{\text{msb}} + 1 = 2$$

$$y_0 = 2^{-2} = 0000.010000000000 = (0.25)_{10}$$
$$y_1 = 0.25(2 - 3 \cdot 0.25) = (0.3125)_{10} = 0000.010100000000$$
$$y_2 = 0.3125(2 - 3 \cdot 0.3125) = (0.33203)_{10} = 0000.010101001111$$
$$\ldots$$

The accuracy will improve with iterations and converge towards $y = (0.33333)_{10}$ or 0000.0101 0101 0101.

Alternate Algorithm

The computation of the reciprocal can be considered as a negative feedback system. The converging portion of the iteration is the transient portion and the steady-state portion is the converged value of the function. Figure 12.3 gives the control block schematic of the reciprocal estimation.

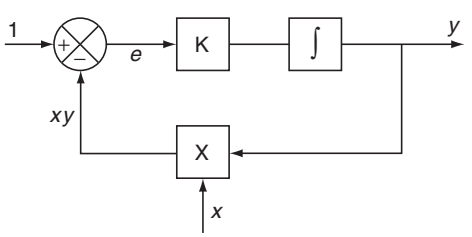


Figure 12.3 Reciprocal algorithm based on control principles.

The function $F(x, y)$ is re-defined as

$$F(x, y) \equiv 1 - x \cdot y = 0 \tag{12.11}$$

From Figure 12.3 and Eq. (12.11), it can be observed that the reference is unity or 1. The output y is multiplied with x and given as feedback. The error is passed through a gain K which is passed through an integrator to obtain the output y. The integrator ensures that the steady-state error is zero in which case $xy = 1$ or the output $y = 1/x$. The integrator can be a digital integrator like the trapezoidal integrator with the gain K tuned to achieve fast transient response.

Square Root

$$\text{Let } y = \sqrt{x} \text{ where } x > 0. \text{ Set}$$

$$F(x, y) \equiv y^2 - x = 0$$

$$\frac{\partial F}{\partial y} = F' = 2y$$

Applying Eq. (12.9), one obtains

$$y_{k+1} = y_k - \left(\frac{y_k^2 - x}{2y_k} \right)$$

Re-arranging the above, the iterative solution for the square root of x is given as

$$y_{k+1} = \frac{1}{2}\left(y_k + \frac{x}{y_k} \right) \text{ where } k = 0, 1, 2, \ldots \tag{12.12}$$

Equation (12.12) is the iterative equation for computing the square root of x. The initial value y_0 is selected as follows:

1. Let x be an n-digit binary number in $\mathbb{D}(2, i, f)$ numeric format. The binary number x can be written as

$$x = 2^m \cdot x_m \text{ where } m \text{ is an integer and } (1/2) \le x_m < 1$$

2. Check the bit number of the MSB of x. Let this be n_{msb}. Then

$$m = n_{\text{msb}} + 1$$

3. Set $y_0 = 2^{\text{trunc}(m/2)}$ in order to have rapid convergence, where trunc(m/2) truncates the number m/2 to the lower integer value.

EXAMPLE 12.7

Consider a 16-bit binary number $x = (5)_{10}$ in $\mathbb{D}(2, 3, 12)$ numeric format:

$$x = 0101.000000000000$$

Here $n_{\text{msb}} = 2$ (corresponding to 2^2). Therefore

$$m = n_{\text{msb}} + 1 = 3$$

$$y_0 = 2^{\text{trunc}(3/2)} = 2^1 = 0010.000000000000 = (2)_{10}$$
$$y_1 = 0.5(2 + 5/2) = (2.25)_{10} = 0010.010000000000$$
$$y_2 = 0.5(2.25 + 5/2.25) = (2.2361)_{10} = 0010.001111000111$$
$$\ldots$$

Alternate Algorithm 1

The square root iteration algorithm given in Eq. (12.12) requires the reciprocal algorithm as the reciprocal of y_k has to be evaluated first before evaluating the square root. This is therefore not a very efficient algorithm in terms of computational speed. Alternately, set

$$F(x, y) \equiv \frac{x}{y^2} - 1 = 0$$

$$\frac{\partial F}{\partial y} = F' = -\frac{2x}{y^3}$$

Applying Eq. (12.9), one obtains

$$y_{k+1} = y_k + \left(\frac{(x / y_k^2) - 1}{2x / y_k^3} \right)$$

Re-arranging the above, the iterative solution for the square root of x is given as

$$y_{k+1} = \frac{y_k}{2}\left(3 - \frac{y_k^2}{x} \right) \text{ where } k = 0, 1, 2, \ldots \tag{12.13}$$

In the above equation, it can be observed that the reciprocal algorithm is still needed for x. However, the reciprocal of x needs to be calculated only once whereas in the case of Eq. (12.12), the reciprocal of y_k needs to be calculated every iteration.

Alternate Algorithm 2

The computation of the square root can be considered as a negative feedback system. Figure 12.4 gives the control block schematic of square root estimation.

The function $F(x, y)$ is re-defined as

$$F(x, y) \equiv x - y^2 = 0 \tag{12.14}$$

From Figure 12.4 and Eq. (12.14), it can be observed that the reference is x. The output y is multiplied with itself and fed back. The error is passed through a gain K which is passed through an integrator to obtain the output y. The integrator ensures that the steady-state error is zero in which case $y^2 = x$ or the output $y = \sqrt{x}$. The integrator can be a digital integrator like the trapezoidal integrator with the gain K tuned to achieve fast transient response.

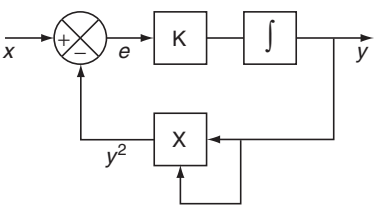


Figure 12.4 | Square root algorithm based on control principles.

Reciprocal of Square Root

Let $y = 1/\sqrt{x}$ where $x > 0$.The function can be re-written as $y = \sqrt{1/x}$. Applying the iteration algorithm for square root as given in Eq. (12.13), one obtains

$$y_{k+1} = \frac{y_k}{2}(3 - x \cdot y_k^2) \quad \text{where } k = 0, 1, 2, \ldots \tag{12.15}$$

The initial value y_0 is selected as follows:

1. Let x be an n-digit binary number in $\mathbb{D}(2, i, f)$ numeric format. The binary number x can be written as

$$x = 2^m \cdot x_m \text{ where } m \text{ is an integer and } (1/2) \le x_m < 1$$

2. Check the bit number of the MSB of x. Let this be n_{msb}. Then

$$m = n_{\text{msb}} + 1$$

3. Set $y_0 = 2^{-\text{trunc}(m/2)}$ in order to have rapid convergence, where trunc(m/2) truncates the number m/2 to the lower integer value.

Extension to Square Root Calculation

The above algorithm can be used to compute the square root of a number x using the reciprocal process.

Let $y = \sqrt{x}$ where $x > 0$. This function can be re-written as $y = \sqrt{x} = x \cdot \sqrt{1/x}$ where $x > 0$. Using Eq. (12.15), $\sqrt{1/x}$ can be computed without any reciprocal operation. After computing $\sqrt{1/x}$, multiply the resulting output value with x to obtain the square root of x.

Hypotenuse

Let a and b be the sides of a right-angled triangle and let h be its hypotenuse. Then by Pythagoras theorem,

$$h = \sqrt{a^2 + b^2} \tag{12.16}$$

Equation (12.16) can be computed using one of the square root algorithms. Alternately, a very fast open-loop algorithm for calculating square roots can be used. It is based on the fact that for a triangle, the hypotenuse h is greater than either of the other two sides a and b. Therefore,

$$h = \max(a, b) + k \cdot \min(a, b) \tag{12.17}$$

where $0 \le k \le 1$. Here max(a, b) is the maximum value among a and b and min(a, b) is the minimum value among a and b. Let

$$x = \max(a, b) \tag{12.18}$$

$$y = \min(a, b) \tag{12.19}$$

Substituting Eqs. (12.18) and (12.19) into Eq. (12.17), one obtains

$$h = x + ky \tag{12.20}$$

From Eq. (12.14), one has

$$h = x + ky \approx \sqrt{x^2 + y^2} \tag{12.21}$$

The error in percent between the approximate algorithm and the actual value obtained by the Pythagoras theorem is

$$e = \frac{\sqrt{x^2 + y^2} - (x + ky)}{\sqrt{x^2 + y^2}} \times 100 \tag{12.22}$$

Using a computer program, x and y can be normalized and k can be varied from 0 to 1. The minimum error obtainable is at $k = 0.34$ where the error is guaranteed to be less than 5.55%. However, from an implementation point of view, $k = 0.5$ is very desirable as scaling by 0.5 can be achieved by simply shifting the data right by one bit. But at $k = 0.5$, it can be guaranteed that error will be less than 11.8%. If this maximum possible error is acceptable, then this would be a very fast hypotenuse computation algorithm.

Polynomial

Most of the digital filters are polynomials. Therefore, evaluating polynomials for a specified value of the parameter x is very common in digital processing applications. Consider a polynomial of degree n. Then,

$$P(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n \tag{12.23}$$

where a_i are real coefficients. If it is required to evaluate the polynomial $P(x)$ at $x = \alpha$, then

$$\begin{aligned} P(\alpha) &= a_0 \alpha^n + a_1 \alpha^{n-1} + \cdots + a_{n-1} \alpha + a_n \\ &= (\ldots(((a_0 \alpha + a_1) \cdot \alpha + a_2) \cdot \alpha + a_3)\alpha + \cdots + a_{n-1})\alpha + a_n \end{aligned} \tag{12.24}$$

The numbers are computed successively as follows:

$$\begin{aligned} b_0 &= a_0 \\ b_1 &= a_1 + b_0 \alpha \\ b_2 &= a_2 + b_1 \alpha \\ &\vdots \\ b_n &= a_n + b_{n-1} \alpha \end{aligned}$$

Then
$P(\alpha) = b_n$

Sine and Cosine

The sine and cosine of an angle are used in many algorithms in power electronics. Therefore, computing sine or cosine of an angle is important. If the digital processor contains ample data memory space, then the best approach from the point of view of speed is to pre-compute the sine values for angles varying from 0° to 360° and store them in the data memory space. Two pointers may be used: one for the sine and the other for the cosine. The two pointers point to the sine table such that they are displaced in angle corresponding to 90°. This means that if the sine pointer points to the memory location corresponding to $\sin(\alpha)$, then the cosine pointer points to the memory location corresponding to $\sin(\alpha + 90)$. Alternately, one may compute the sine and cosine values by using the series expansion method based on the following discussion.

The sine and cosine may be calculated by their series expansions. Let α and β be the angles such that $0 \le \alpha \le \pi/2$ and $\beta = (\pi/2) - \alpha$. If $0 \le \alpha \le \pi/4$, then $\pi/4 \le \beta \le \pi/2$. The sine and the cosine are given as

$$\cos\beta = \sin\alpha = \sum_{n=0}^{\infty} (-1)^n \frac{\alpha^{2n+1}}{(2n+1)!} \tag{12.25}$$

If $\pi/4 \le \alpha \le \pi/2$ then $0 \le \beta \le \pi/4$. The sine and the cosine are given as

$$\sin\alpha = \cos\beta = \sum_{n=0}^{\infty} (-1)^n \frac{\beta^{2n}}{(2n)!} \tag{12.26}$$

Equation (12.25) can be computed by the summation of

$$\cos\beta = \sin\alpha = s_1 + s_2 + \cdots + s_k + s_{k+1} + \cdots \tag{12.27}$$

where s_k are found by the following relationships:

$$s_1 = \alpha$$

$$s_{k+1} = -\frac{\alpha^2}{2k(2k+1)} s_k \quad (k = 1, 2, \ldots)$$

The summation process can be terminated when $|s_k| \le \in$, where $\in$ is a small number indicating the residual error.

Similarly, Eq. (12.26) can also be computed by the summation of

$$\sin\alpha = \cos\beta = c_1 + c_2 + \cdots + c_k + c_{k+1} + \cdots \tag{12.28}$$

where c_k are found by the following relationships:

$$c_1 = 1$$

$$c_{k+1} = -\frac{\beta^2}{2k(2k-1)} c_k \qquad (k = 1, 2, \ldots)$$

The summation process can be terminated when $|c_k| \le \in$, where $\in$ is a small number indicating the residual error.

Exponential

The exponential function e^x can be expanded in series form as

$$e^x = 1 + x + \frac{x^2}{2!} + \quad + \frac{x^n}{n!} + \tag{12.29}$$

where $-\infty < x < \infty$. The number x can be written as

$$x = I + F$$

where I is the integer part and F is the fractional part wherein $0 \le F < 1$. Therefore, e^k can be expressed as

$$e^k = e^I \cdot e^F \tag{12.30}$$

In Eq. (12.30), the first factor is expressed as

$$e^I = e \cdot e \cdot e \cdots e \quad I \text{ times if } I \geq 0$$

or

$$e^I = \frac{1}{e} \cdot \frac{1}{e} \cdot \frac{1}{e} \cdots \frac{1}{e} \quad I \text{ times if } I < 0$$

where $e = 2.7182818\ldots$ and $1/e = 0.3678794\ldots$.
In Eq. (12.30), the second factor is expressed as

$$e^F = 1 + F + \frac{F^2}{2!} + \frac{F^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{F^n}{n!} \qquad \text{where } F < 0$$

This can be computed by summation of the series

$$e^F = u_1 + u_2 + \cdots + u_k + u_{k+1} + \cdots$$

where

$$u_0 = 1$$

$$u_{k+1} = \frac{F}{k+1} u_{\mathrm{k}} \quad (k = 1, 2, \ldots)$$

The summation process can be terminated when $|u_k| \leq \in$, where $\in$ is a small number indicating the residual error.

Logarithm

The series expansion of the natural logarithm of numbers is given as

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{k-1} \frac{x^k}{k} + \cdots \tag{12.31}$$

where $-1 < x \leq 1$ or $0 < 1 + x \leq 2$. Equation (12.31) is very limited in its application as the range of numbers x is small and for values of x whose absolute value is close to unity, the convergence is very slow. To solve this problem of range and convergence, the following modification to the series is performed. Substituting x with $-x$ in Eq. (12.31), one obtains

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots - \frac{x^k}{k} - \cdots \tag{12.32}$$

Subtracting Eq. (12.32) from Eq. (12.31), one obtains

$$\ln\left(\frac{1-x}{1+x}\right) = -2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\right) \tag{12.33}$$

Set $z = (1-x)/(1+x)$. This means $x = (1-z)/(1+z)$. Substituting the value of x in Eq. (12.33), one has

$$\ln z = -2\left[\left(\frac{1-z}{1+z}\right) + \frac{1}{3}\left(\frac{1-z}{1+z}\right)^3 + \frac{1}{5}\left(\frac{1-z}{1+z}\right)^5 + \cdots\right] \tag{12.34}$$

For x in the range $-1 < x \le 1$, z is in the range $0 < z < \infty$. Equation (12.34) can now handle the full range of the positive real numbers.
Let x be an n-digit binary number in $\mathbb{D}(2, i, f)$ numeric format. The binary number x can be written as

$$x = 2^m \cdot x_m \quad \text{where } m \text{ is an integer and } \frac{1}{2} \le x_m < 1$$

From Eq. (12.34),

$$\ln x = m \ln 2 + m \ln x_m$$

$$\ln x = m \ln 2 - 2\left[\left(\frac{1-x_m}{1+x_m}\right) + \frac{1}{3}\left(\frac{1-x_m}{1+x_m}\right)^3 + \frac{1}{5}\left(\frac{1-x_m}{1+x_m}\right)^5 + \cdots\right] \tag{12.35}$$

Let $\alpha = (1-x_m)/(1+x_m)$ then $0 < \alpha \le 1/3$ and

$$\ln x = m \ln 2 - 2\left(\alpha + \frac{\alpha^3}{3} + \frac{\alpha^5}{5} + \cdots + \frac{\alpha^{2k-1}}{2k-1} + \cdots\right)$$

This equation can be computed by series summation as

$$\ln x = m \ln 2 - 2\left(u_1 + u_2 + u_3 + \cdots + u_k + \cdots\right)$$

where
ln2 = 0.693147…

$$u_1 = \alpha$$

$$u_{k+1} = \frac{(2k-1)}{(2k+1)} u_k \qquad (k = 1, 2, \ldots)$$

The summation process can be terminated when $|u_k| \le \in$, where $\in$ is a small number indicating the residual error.

Implementation Examples

This section will discuss some application examples that are common in power electronic systems. These application examples are discussed in an algorithmic manner indicating the equations and the sequence in which the equations are arranged for digital computation. The arithmetic operations for computing the equations are based on the algorithms discussed in the previous sections.

PI Controller

The PI controller is discussed in detail in Chapter 10. The block schematic of the PI controller is shown in Figure 12.5. It consists of a comparator that compares the reference x_{ref} and the feedback signal x_f. The error e_{pi} is fed to the proportional and the integral parts of the controller as shown in the figure. The integrator of the controller is a limited integrator to prevent windup. The output of the PI part v_{cont} is added with a feed-forward signal if present and passed through a clamper to limit the controller output V_c to within the band bounded by V_{max} and V_{min}. Together with the Figure 12.5, the algorithm given below is self-explanatory.

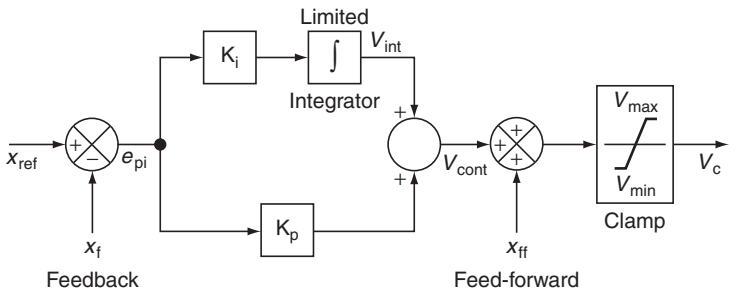


Figure 12.5 | Implementational block schematic of a PI controller.

```
[vc] = pi_controller(xref, xf, xff)
Local Variables:
epi, vint, vcont
Constants:

//*** PI regulators parameters
Ki, Kp, Kclamp

//*** vc clamp limits
Vmin, Vmax
{
   epi = xref – xf;
   //integrator portion
       if (vc < Vmax)& (vc > Vmin)
       Kclamp=1;
       else
       Kclamp=0;
       end
   vint = vint + Kclamp*Ki*epi; //integrator output with anti-windup
   //proportional plus integral portion
   vcont = vint + Kp*epi;
   //add feedforward term and clamp output
   if (vcont + xff)>= Vmax
       vc = Vmax;
   elseif (vcont + xff) <= Vmin
          vc = Vmin;
   else
       vc = vcont + xff;
   end
}
```

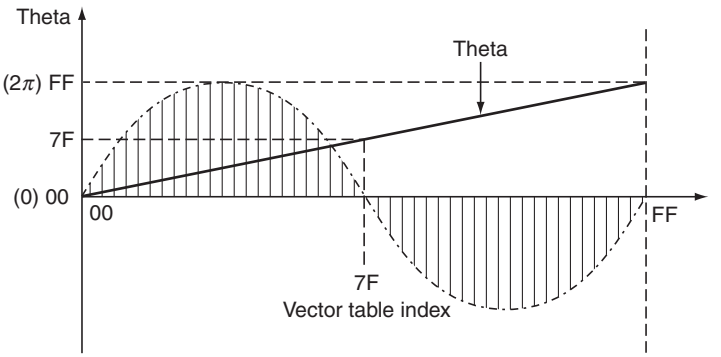


Figure 12.6 | Sine table and the angle relationship.

Sine and Cosine

The sine and cosine of an angle are obtained here by browsing through a pre-computed sine table. Assume that the data is 16-bit data, but say only 256 data words of memory storage are available in the digital processor. To create a sine table, generate sine values from 0 to 2π in $\mathbb{D}(2, 0, 15)$ with one bit being used for sign bit for the 2's complement number. As the sine values are less than unity, no integer bits are required. From the generated table (containing 2^{16} 16-bit data), extract 256 16-bit data by selecting only those wherein the most significant 8 bits are *distinct*. This would form the sine table. The starting address of the sine table (*sin_table) is also passed to the function along with the 16-bit angle data to compute the sines and cosines.

The algorithm is simple and can be understood clearly from Figure 12.6. The angle Theta can take values from 0 to 2π, that is from $(0000)_{16}$ to $(FFFF)_{16}$. The data table may be considered as a vector wherein the length varies from 0 to 255, that is, $(0000)_{16}$ to $(00FF)_{16}$. In order to map the angle to the data table vector, an angle of Theta = $(00xx)_{16}$ maps to the sine table vector element $(0000)_{16}$ and the angle of Theta = $(FFxx)_{16}$ maps to the sine table vector element $(00FF)_{16}$. This is a many-to-one mapping where "*x*" is a "don't care" term. The starting address of the sine table, "*sin_table", is used to offset the sine table vector index. For obtaining the cosine of the angle Theta, an additional offset corresponding to one-fourth, the length of the sine table has to be included in the index to obtain a 90° phase shift.

```
[sinTheta, cosTheta] = sincos(Theta, *sin_table)
Local Variables:
index, sine_table
{
   //obtaining sine of the angle
   index=MSByte(Theta);//Most signifcant byte of Theta is index of
                       //the sine table vector
   index = index + *sin_table;//offset index with starting address of
                       // sine table
   sinTheta = sine_table(index);
                       //obtaining cosine of the angle
```

```
   index =MSByte(Theta);//Most signifcant byte of Theta is index of
                  //the sine table vector
   index = index + *sin_table + (0040)16;//offset index with starting
                                 //address of sine table.
                              //(0040)16 additional offset corresponds
                              //to a 90° phase shift.
   cosTheta = sine_table(index);
}
```

Pulse-Width Modulation (PWM)

The PWM has been discussed in Chapter 6. However, this section presents an algorithm for implementing the single-phase PWM. The input to the algorithm is the reference signal, V_{ref}. The input reference signal is compared notionally with a triangular waveshape. Due to the linear structure of a triangular waveform, the pulse widths are proportional to the voltages at the instant of comparison with the triangular carrier. Therefore, if V_r is the voltage of the reference signal at the instant of comparison, then

$$t_r = \left(\frac{V_r}{V_{dc}}\right)\frac{T_s}{2} \tag{12.36}$$

where V_{dc} is the maximum attainable peak voltage of the phase waveforms that corresponds to modulation index of 1 and T_s is the switching time period. The value of V_r may be positive or negative depending on the nature of the reference waveshape. If the reference waveshape is a DC signal then V_r is always positive and hence t_r is also positive. If the reference waveshape is an AC signal like in single-phase sinusoidal PWM, then V_r can be either positive or negative. Both these cases will be discussed.

DC Reference

Figure 12.7 shows the DC reference wave compared with a triangular carrier. This is generally encountered in DC–DC converters. The reference waveform is sampled at an arbitrary comparison instant. The expanded time scale at the sample instant shows the comparison that spans a carrier period or switching period T_s. At the comparison instant the sampled value is V_r. During the period T_s, the sampled value is compared with the triangular waveform as shown:

$$\frac{T_0}{2} = \frac{T_s}{2} - t_r \quad \text{or} \quad T_0 = T_s - 2t_r \tag{12.37}$$

where T_0 is the time when the PWM output is zero.

Over-Modulation The maximum value of the converter output voltage is limited by the DC bus voltage. To maintain linearity, the reference amplitude is maintained within the DC bus voltage. However, during transients, the reference amplitudes may go beyond the DC bus. Such a situation is called over-modulation. During over-modulation, t_r becomes greater than T_s and T_0 becomes negative. To compensate for this, the times are multiplied by T_s/t_r to clamp the average output to the DC bus. This is a simple over-modulation technique sufficient for many applications.

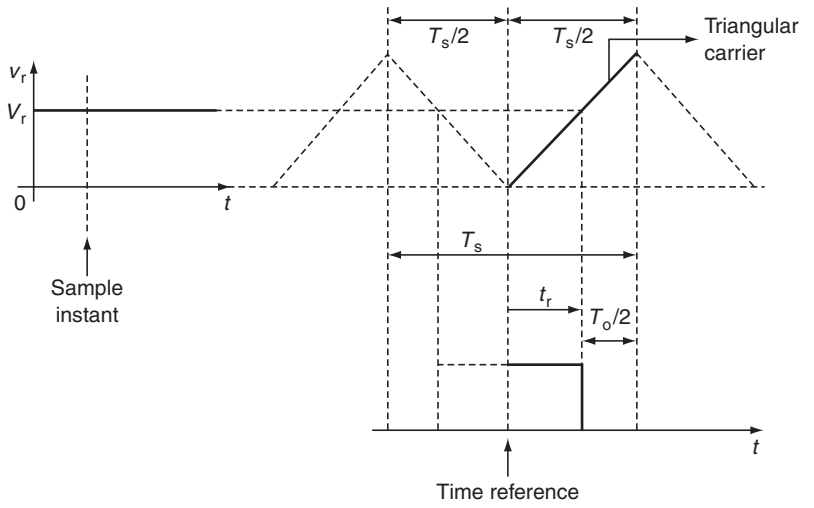


Figure 12.7 | PWM generation with DC reference.

The generalized algorithm for the PWM generation with DC reference is as follows:

```
[tr] = pwm(Vr)
Local Variables:
tr, toffset, Ts, Vdc, T0
{
   tr=Ts*Vr/Vdc/2;
   T0 = Ts - 2*tr;
//Overmodulation
   if (T0 < 0),
   tr = Ts/2;
   end
}
```

The value of t_r obtained from this algorithm should be pushed into a compare register of the digital processor that makes comparison with the triangular waveform. The triangular carrier waveform is generated by programming one of the internal general-purpose timers of the digital processor as an up–down counter to emulate a triangle waveform having a period T_s. The output of the compare registers will produce the PWM waveforms.

AC Reference

Figure 12.8 shows an AC reference waveform. This is generally encountered in single-phase inverter applications. The reference waveform v_r and an inverted v_r as shown in Figure 12.8 are used to generate the PWM for two-level two-arm inverters. The reference v_r is applied to arm "*a*" of the inverter and is denoted as v_{ra} and the inverted reference is applied to arm "*b*" of the inverter and is denoted as v_{rb}. It is required to obtain the ON-times t_{aon} and t_{bon} for the two inverter arms. If V_a and V_b are the sampled voltages of v_{ra} and v_{rb} at a specific sample instant, then

$$t_a = \left(\frac{V_a}{V_{dc}}\right) T_s \tag{12.38}$$

$$t_b = \left(\frac{V_b}{V_{dc}}\right) T_s \tag{12.39}$$

where V_{dc} is the maximum attainable peak voltage of the phase waveforms that corresponds to modulation index of 1 and T_s is the switching time period. The value $t_{max} = \max(t_a, t_b)$ gives the maximum value among t_a and t_b; $t_{min} = \min(t_a, t_b)$ gives the minimum value among t_a and t_b.

Figure 12.8 shows the steps involved in generating the PWM waveform. The two outputs of phase reference waveforms are sampled at an arbitrary comparison instant. This comparison instant spans a period T_s. The comparison instant is shown expanded wherein the two-phase sampled values are V_a and V_b, as shown. During the period T_s, the sampled values are compared with the triangular waveform as shown. At that instant, the waveshape having the largest voltage will have a time t_{max}. In the figure, V_a is shown to have a maximum value and therefore $t_{max} = t_a$ in this case. As V_b is the minimum value at this instant, $t_{min} = t_b$.

From Figure 12.8, it can be observed that t_{max} which is measured right from the central reference is positive. t_{min} is measured left from the central reference and is a negative number as V_b is negative during the sampled instant. The time value t_{min} cannot be negative. Therefore an offset has to be provided to the central reference line so that the time values are positive. The new time reference is shifted from the central line to the line at the beginning of the T_s interval as shown in the figure. The shift from the central line to the new reference line is the time offset which is given as

$$t_{offset} = |t_{min}| + \frac{T_0}{2}$$

$$t_{offset} = -t_{min} + \frac{T_0}{2} \tag{12.40}$$

where T_0 is the time duration when the PWM output across the inverter bridge is zero in a T_s period.

Over-Modulation The maximum value of the inverter output voltage is limited by the DC bus voltage. The reference amplitude is maintained within the DC bus voltage. However, during transients, the peak reference amplitude may go beyond the DC bus value. Such a situation is called over-modulation. During over-modulation, $t_{max} - t_{min}$ becomes greater than T_s and T_0 becomes negative. To compensate for this, the ON-times are scaled to clamp the peak output to the DC bus. This is a simple over-modulation technique sufficient for many applications.

The generalized algorithm for the PWM generation for AC reference is as follows.

```
[taon, tbon] = spwm(Va, Vb)
Local Variables:
tmin, tmax, toffset, Ts, Vdc, Teff, T0
{
   //fnd ta, tb
   ta=Ts*Va/Vdc;
   tb=Ts*Vb/Vdc;
   //fnd tmax  and tmin
       tmax  = ta;
```

```
      if (tb > tmax)
          tmax = tb;
      end
      tmin = ta;
      if (tb < tmin)
          tmin = tb;
      end
Teff = tmax - tmin;
T0 = Ts - Teff;
Toffset = (T0/2) - tmin; //offset time from the origin.
//the on-time for each of the PWM phases is given by:
taon = ta + toffset;
tbon = tb + toffset;
//Overmodulation
   if (T0 < 0),
   toffset = -tmin*Ts/Teff;
   taon = ta*Ts/Teff + toffset;
   tbon = tb*Ts/Teff + toffset;
   end
}
```

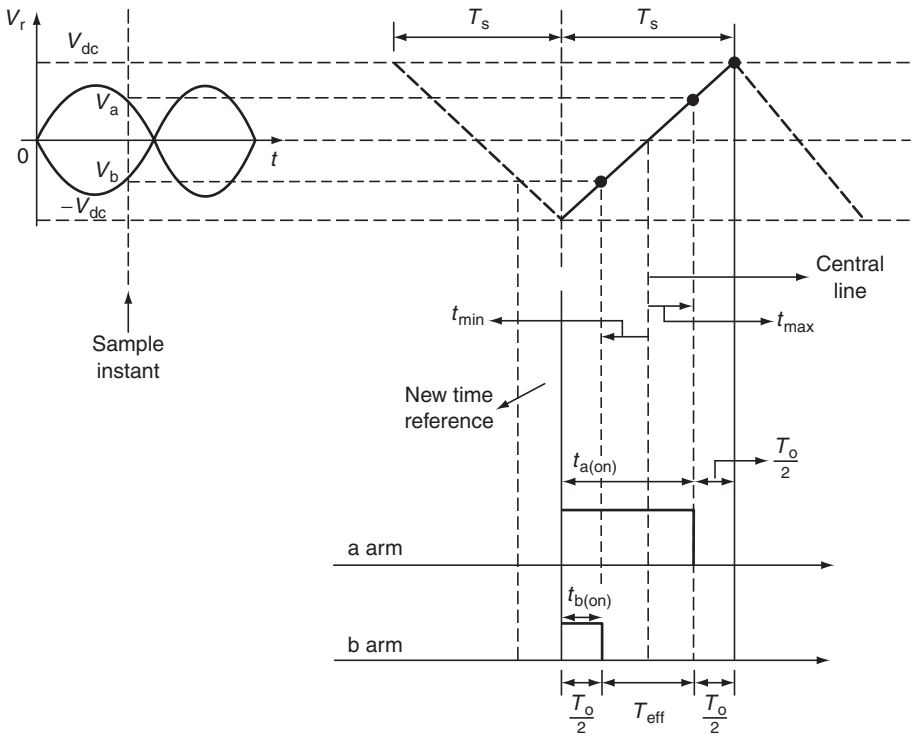


Figure 12.8 | PWM generation with AC reference.

The values t_{aon} and t_{bon} should be pushed into a compare register of the digital processor that makes comparison with a triangular waveform. The triangular carrier waveform is generated by programming one of the internal general-purpose timers of the digital processor as an up–down counter to emulate a triangle waveform having a period $2T_s$. The output of the compare registers will produce the PWM waveforms.

Space Vector PWM

The space vector PWM and its generation have been discussed in Chapter 6. However, this section presents an algorithm for implementing the space vector PWM that is much simpler and faster. The inputs to the algorithm are the three-phase reference sine waves, v_{aref}, v_{bref} and v_{cref}. The input reference waveshapes are compared notionally with a triangular waveshape. Due to the linear structure of a triangular waveform, the pulse widths are proportional to the voltages at the instant of comparison with the triangular carrier. Therefore, if V_a, V_b and V_c are the voltages of the three-phase waveforms at the instant of comparison with the triangular wave, then

$$t_a = \left(\frac{V_a}{V_{dc}}\right) T_s \tag{12.41}$$

$$t_b = \left(\frac{V_b}{V_{dc}}\right) T_s \tag{12.42}$$

$$t_c = \left(\frac{V_c}{V_{dc}}\right) T_s \tag{12.43}$$

where V_{dc} is the maximum attainable peak voltage of the phase waveforms that corresponds to modulation index of 1 and T_s is the switching time period.

Figure 12.9 shows the three-phase waveform and the steps in the generation of the corresponding timing for each of the PWM time intervals T_0, T_1 and T_2, respectively, where T_0 is the time spent in zero space vector, T_1 is the time spent in the primary space vector of the sector and T_2 is the time spent in the secondary space vector of the sector. The notations T_0, T_1 and T_2 connote the same meanings as explained in Chapter 6 in the discussion of space vector PWM. The value $t_{max} = \max(t_a, t_b, t_c)$ gives the maximum value among t_a, t_b and t_c and $t_{min} = \min(t_a, t_b, t_c)$ gives the minimum value among t_a, t_b and t_c.

Figure 12.9 shows the steps involved in generating the PWM waveform. The three-phase reference waveform is sampled at an arbitrary comparison instant. This comparison instant spans a period T_s. The comparison instant is shown expanded wherein the three-phase sampled values are V_a, V_b and V_c as shown. During the period T_s, the sampled values are compared with the triangular waveform as shown. At that instant, the waveshape having the largest voltage will have a time t_{max}. In the figure, V_b is shown to have a maximum value and therefore $t_{max} = t_b$ in this case. As V_c is the minimum value at this instant, $t_{min} = t_c$. Also, V_a is in between V_b and V_c and therefore, $t_{mid} = t_a$.

From Figure 12.9, it can be observed that t_{max} which is measured right from the central reference is positive. Both t_{min} and t_{mid} are measured left from the central reference and are negative as V_a and V_c are negative during the sampled instant. The time values t_{min} and t_{mid} cannot be negative values. Therefore an offset has to be provided to the central reference line so that the time values are positive. The new time reference is shifted from the central line to the line at the beginning of the T_s interval as shown in the figure. The shift from the central line to the new reference line is the time offset and is given as

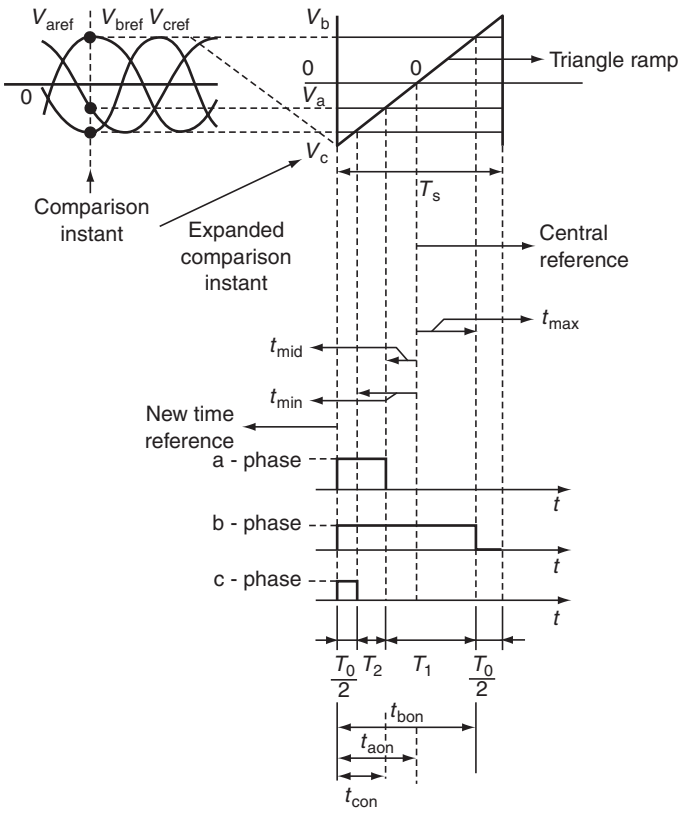


Figure 12.9 | Generation of SPWM waveform.

$$t_{offset} = |t_{min}| + \frac{T_0}{2}$$

$$t_{offset} = -t_{min} + \frac{T_0}{2} \tag{12.44}$$

where T_0 is the time duration when the system is in zero space vector state.

The time interval $t_{max} - t_{min}$ is denoted as T_1 and is the time interval for which the system is at the space vector corresponding to the maximum phase value which is phase *b* in this case. The time interval $t_{mid} - t_{min}$ is the time interval for which the system is at the space vector corresponding to the middle phase value which is phase *a* phase in this case. The values of T_0, T_1 and T_2 are obtained from the instantaneous values of the waveform. For every alternate cycle, one must take the mirror image of this sequence as the triangular waveform slope alternates between positive and negative slopes. For the case of Figure 12.9, phase *b* has the maximum voltage and therefore it is ON till time t_{bon} ($t_{max} + t_{offset}$) from the new reference. The phase *a* has the in-between voltage and therefore, it is ON till time t_{aon} [$(T_0/2) + t_{mid} - t_{min}$] from the new reference. Phase *c* has the minimum voltage and therefore, it is ON for only t_{con} ($T_0/2$) time.

Over-Modulation During over-modulation, $t_{max} - t_{min}$ becomes greater than T_s and T_0 becomes negative. To compensate for this, the times are multiplied by $T_s/(t_{max} - t_{min})$ such that the resulting space vector is clamped to the hexagon boundary.

The generalized algorithm for the PWM generation is as follows.

```
[taon, tbon, tcon] = svpwm(Va, Vb, Vc)
Local Variables:
tmin, tmax, toffset, Ts, Vdc, Teff, T0
{
   //fnd ta, tb, tc
   ta=Ts*Va/Vdc;
   tb=Ts*Vb/Vdc;
   tc=Ts*Vc/Vdc;
   //fnd tmax  and tmin
       tmax = ta;
       if (tb > tmax)
           tmax = tb;
       elseif (tc > tmax)
           tmax = tc;
       end
       tmin = ta;
       if (tb < tmin)
           tmin = tb;
       elseif (tc < tmin)
           tmin = tc;
       end
Teff = tmax - tmin;
T0 = Ts - Teff;
Toffset = (T0/2) - tmin; //offset time.
//the on-time for each of the PWM phases is given by:
taon = ta + toffset;
tbon = tb + toffset;
tcon = tc + toffset;
//Overmodulation
   if (T0 < 0),
   toffset = -tmin*Ts/Teff;
   taon = ta*Ts/Teff + toffset;
   tbon = tb*Ts/Teff + toffset;
   tcon = tc*Ts/Teff + toffset;
   end
}
```

Here t_{aon}, t_{bon} and t_{con} should be pushed into compare registers of the digital processor that makes comparison with the triangular waveform. The triangular carrier waveform can be generated by programming one of the internal general-purpose timers of the digital processor as an up–down counter to emulate a triangle waveform having a period $2T_s$. The output of the compare registers will produce the PWM waveforms.

Over-Modulation

The maximum value of the inverter output voltage is limited by the DC bus voltage. To maintain linearity, the reference space vector amplitudes are maintained within the space vector hexagon. However, during transients, the reference space vector amplitudes may go beyond the hexagon. Such a situation is called over-modulation. Such reference space vector demand must be detected and appropriately limited. There are many strategies existing in the literature to treat the reference space vectors during over-modulation. However, the over-modulation strategies fall into the following three categories: (a) circular boundary limit, (b) hexagonal boundary limit and (c) expanding switching period.

Circular Boundary Limit

In this method, the largest circle that can be inscribed within the space vector hexagon is considered as the boundary of the space vector. Therefore, in over-modulation situations, the reference space vector amplitude is limited to the circular boundary as shown in Figure 12.10.

Referring to Figure 12.10, V_r^* is the reference space vector that is demanded. As V_r^* is beyond the maximum circular boundary limit and is limited to the circular boundary V_{rm}. The circular boundary limit is given by

$$V_{rm} = V_{dc} \cdot \cos 30 = \frac{\sqrt{3}}{2} V_{dc} \cdot$$

The space vector V_r^* can be considered to be resolved into $V_{s\alpha}{}^*$ and $V_{s\beta}{}^*$ as indicated in Figure 12.10.

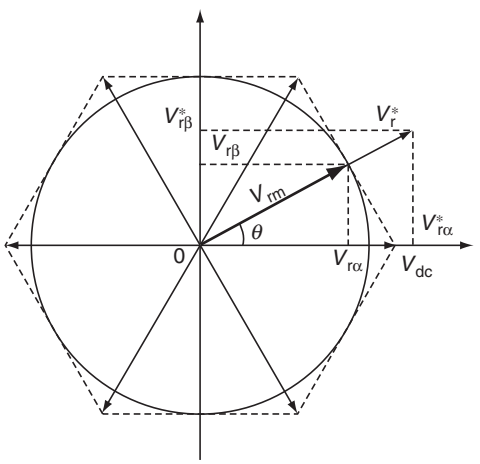


Figure 12.10 | Circular boundary limit within the space vector hexagon.

Thus,

$$V_r^* = \sqrt{V_{r\alpha}^{*2} + V_{r\beta}^{*2}}$$

If $V_r^* = \sqrt{V_{r\alpha}^{*2} + V_{r\beta}^{*2}} > V_{rm}$, then V_r^* is limited to V_{rm} such that

$$V_{r\alpha} = \frac{V_{rm}}{\sqrt{V_{r\alpha}^{*2} + V_{r\beta}^{*2}}} \cdot V_{r\alpha}^*$$

$$V_{r\beta} = \frac{V_{rm}}{\sqrt{V_{r\alpha}^{*2} + V_{r\beta}^{*2}}} \cdot V_{r\beta}^*$$

$$V_r = \sqrt{V_{r\alpha}^{2} + V_{r\beta}^{2}}$$

Thus it can be observed that in over-modulation situation, this method applies the largest possible voltage space vector that is oriented in the same direction as the demanded reference voltage space vector and is on the maximum circular locus that can be inscribed within the space vector hexagon. The inverter switches cannot respond to small pulse widths near the state vectors. Therefore, if it is required to eliminate pulse dropping then the circular locus may be accordingly modified to a near circular locus to ensure that the pulse widths near the state vectors do not become very small.

Hexagonal Boundary Limit

The circular boundary limit method requires the square root and reciprocal algorithms to be used. The hexagonal boundary limit method does not require the square root or the reciprocal functions. Therefore, this method is more processor-friendly than the circular boundary limit method. The space vector PWM implementation discussed in the previous sub-section actually uses this hexagonal boundary limit.
Here, during over-modulation

$$T_1 + T_2 > T_s$$

$$T_{eff} = (t_{max} - t_{min}) > T_s$$

In this case the calculated times are scaled with $T_s/(T_1 + T_2)$ or T_s/T_{eff} to produce a realizable vector on the space vector hexagon as shown in Figure 12.11. The realized space vector is again oriented in the

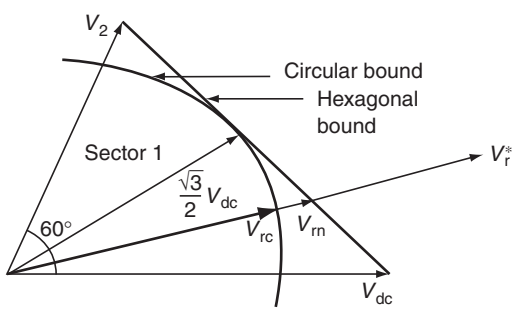


Figure 12.11 | Hexagonal boundary limit.

same direction as the demanded reference space vector but it now lies on the hexagonal boundary rather than the circular the boundary. This algorithm is simple to implement on a digital processor but the resulting line voltages contains low frequency distortions. However, in this method the voltage vectors obtainable near the state vectors of the sector are larger than that of the circular boundary method as shown in Figure 12.11.

Referring to Figure 12.11, V_r^* is the demanded space vector, V_{rc} is the resulting space vector on using the circular boundary limit method and V_{rn} is the resulting space vector on using the hexagonal boundary limit method.

Expanding Switching Period

In the previous two methods, the switching period T_s is held constant. To better utilize the inverter especially for high-power drives, the inverter drive pulse should seamlessly move from the high-frequency carrier PWM to the low-frequency six-step waveform as described in Chapter 6. For this the switching period T_s is gradually expanded till six-step operation results. There are numerous algorithms in the literature that discuss the smooth transition from PWM to six-step each with their advantages and disadvantages. A simple method of expanding the switching period is as follows.

During over-modulation situation, $T_1 + T_2 > T_s$ or $T_{eff} > T_s$. For a demanded reference space vector, calculate T_1 and T_2 for the active state vectors. If $T_1 + T_2 > T_s$ then $T_0 < 0$. Set $T_s = T_1 + T_2$. This implies that the period register of the processor's general-purpose timer that is being used as an up–down counter to emulate a triangular carrier is changed. With this changed period, $T_0 = 0$ and the inverter system spends T_1 period in one active state vector of the sector and the remaining period (T_2) is spent on the other active state vector of the sector. Progressively, if the demanded reference space vector becomes larger and larger, the carrier period for those space vectors is progressively increased till six-step operation is reached.

CONCLUDING REMARKS

This chapter on discrete computational essentials is a rather important one. The implementation of the power electronic circuits can be completed only if the control algorithms are successfully incorporated within a digital processor. The concept of the real number is much different when viewed from within the digital domain due to the two primary limitations of the discrete real numbers: (a) range and (b) resolution. In this regard, it becomes very important to select a numeric format for a particular application. Simple arithmetic processes like addition, subtraction and multiplication can give wrong results if the numeric format is not properly chosen.

Normalization of the variables within the digital domain is another important consideration that needs to be addressed before implementing an algorithm. The choice of the base values of variables for normalization is another classic case of trade-off between precision on one hand which has a bearing on the minimum value of the variables and range on the other hand to accommodate short-term overflows like over-voltage and over-currents in practical applications.

This chapter discusses various arithmetic algorithms that are essential for executing the linear and non-linear models of the power electronic systems for the purposes of control. The framework of arithmetic computation is based on iterative algorithms in order to make the computations repeatable, simple and self-terminating. The next section provides few tutorial exercises that will give the necessary practice for implementing even complex applications.

TUTORIAL EXERCISES

1. Capture and convert analog voltage to discrete domain

Mode of implementation:

a. Microcontroller or DSP Kit

Tasks for study:

(a) Connect a potentiometer to a 5 V supply and connect the tap point to the ADC connector input. (Keep the tap value to less than ADC pin voltage rating).

(b) Write an interrupt service routine (ISR) to acquire the analog signal and observe the acquired value by transferring the value in the ADC FIFO register to a variable. Observe the variable value in the WATCH window of the debugger.

(c) Calculate the *a priori* value that should be obtained in the variable register and compare with the result displayed in the debugger WATCH window.

(d) Now transfer the variable to a DAC ports and observe on the oscilloscope.

(e) In order to compute the time taken by the above interrupt service routine.

i. At the beginning of the ISR make an I/O port high.

ii. At the end of the ISR make that I/O port low.

iii. Observe the waveform of that I/O port on the scope.

iv. Measure the ON-duration of the pulses to obtain the ISR compute time.

2. Computation of a space vector using the open-loop hypotenuse algorithm.

Mode of implementation:

a. Microcontroller or DSP Kit

Tasks for study:

(a) Connect a potentiometer to a 5 V supply and connect the tap point to the ADC connector input. (Keep the tap value less than the ADC input voltage rating).

(b) Acquire the signal and let that represent V_m in $V_m \sin\omega t$.

(c) Include a sine table in the data memory space.

(d) Using pointers that are spaced an equivalent 90° apart, read out the sine and cosine values from the table and multiply with V_m. The read-out rate (frequency of the waveforms) can be fixed by the timer that interrupts the interrupt service routine (ISR).

(e) This will result in two waveforms: $V_m \sin\omega t$ and $V_m \cos\omega t$. Let them be named as V_α and V_β, respectively.

(f) Now apply the following algorithm,

i. Compare V_α and V_β.

ii. Evaluate Larger of $[V_\alpha, V_\beta] + 0.5$ *times Smaller of* $[V_\alpha, V_\beta]$.

iii. Comment on the result.

(g) In MATLAB compute the maximum error that can occur between the following two algorithms.

i. sqrt($V^2_\alpha + V^2_\beta$).

ii. *Larger of* $[V_\alpha, V_\beta] + 0.5$ *times Smaller of* $[V_\alpha, V_\beta]$.

iii. Compare with results of step 6 and comment on the results

3. Normalization and Scaling.

Mode of implementation:

a. Microcontroller or DSP Kit.

Tasks for study:

(a) Acquire the signal and let that represent V_m in $V_m \sin\omega t$.

(b) Include a sine table in the data memory space.

(c) Use the following base values for normalization and compute the scaling factor

i. $V_{base} = V_{max}$

ii. $V_{base} = 0.5 V_{max}$

iii. $V_{base} = 0.25 V_{max}$

where V_{max} is the maximum voltage that can be applied to the ADC inputs.

(d) Convert the signal V_m and the resulting sine signal into the $D(2, 3, 12)$ format and observe the sine wave on the oscilloscope.

4. Computation of reciprocal: To compute $y = 1/x$ where $x > 0$

Mode of implementation:

a. Microcontroller or DSP Kit

Tasks for study:

(a) Algorithm 1:

i. $y_{k+1} = y_k(2 - xy_k)$

ii. if $x = 2^m x_1$ and $0.5 \le x_1 < 1$, then $y_0 = 2^{-m}$

iii. Evaluate the above algorithm for speed of convergence in the $D(2, 7, 8)$ format for a given value of x in $D(2, 7, 8)$ format.

(b) Algorithm 2:

i. $y = 1/x$ can be modeled as $(1 - xy) = 0$ or $(1 - xy) = \varepsilon$ where ε is the error.

ii. A typical control system can be used wherein the reference is the value '1' and the feedback is $x \times y$. The error is fed to an integrator through a gain K. An integral control is used to obtain zero error.

iii. Simulate the above first in SIMULINK.

iv. Then implement the same in the DSP using $D(2, 7, 8)$ format.

v. Find out the value of K which results in the fastest convergence to stability.

(c) Compare the speeds of convergence of the above two algorithms.

5. Computation of square roots: to compute $y = \sqrt{x}$ where $x > 0$

Mode of implementation:

a. Microcontroller or DSP Kit

Tasks for study:

(a) Algorithm 1:

i. $y_{k+1} = \frac{1}{2}\left(y_k + \frac{x}{y_k}\right)$

ii. if $x = 2^m x_1$ and $0.5 \le x_1 < 1$, then $y_0 = 2^{\text{truncate}(m/2)}$

iii. Evaluate the above algorithm for speed of convergence in the $D(2, 7, 8)$ format for a given value of x in $D(2, 7, 8)$ format.

(b) Algorithm 2:

i. $y_{k+1} = \frac{y_k}{2}\left(3 - \frac{y_k^2}{x}\right)$

ii. if $x = 2^m x_1$ and $0.5 \le x_1 < 1$, then $y_0 = 2^{\text{truncate}(m/2)}$

iii. Evaluate the above algorithm for speed of convergence in the $D(2, 7, 8)$ format for a given value of x in $D(2, 7, 8)$ format.

(c) Algorithm 3:

i. $y = \text{sqrt}(x)$ can be modeled as $(x - y \cdot y) = 0$ or $(x - y^2) = \varepsilon$ where ε is the error.

ii. A typical control system can be used wherein the reference is the value 'x' and the feedback is y^*y. The error is fed to an integrator through a gain K. An integral control is used to obtain zero error.

iii. Simulate the above first in SIMULINK.

iv. Then implement the same in the DSP using $D(2, 7, 8)$ format.

v. Find out the value of K which results in the fastest convergence to stability.

(d) Compare the speeds of convergence of the above algorithms using the same initial starting values.

6. Implement the following discrete transfer function in a digital processor,

$$H(z) = \frac{1 - e^{-\tau T}}{\tau\left(z - e^{-\tau T}\right)}$$

where T is the sampling time period in seconds and τ is any positive constant in $D(2, 3, 12)$ format.

Mode of implementation:

a. Microcontroller or DSP Kit

Tasks for study:

(a) Apply a very low-frequency square wave as an input to this function and observe the output on the oscilloscope.

(b) Comment on the nature of the output.

(c) Vary τ and observe the output. Comment on the effect of τ on the output.

(d) Increase the frequency of the input square wave gradually. What happens to the output?

FILL IN THE BLANKS

1. Natural real numbers have infinite _____ and infinite _____.

2. In discrete real number finite range results in _____ error.

3. Discrete real number has finite _____ apart from finite range.

4. The relative density is _____ in the floating-point number system.

5. In the fixed-point system, one must keep track of the position of the _____ point during arithmetic operations.

6. There is a limitation in the floating-point system in representing the number _____ in the normalized format.

7. The floating-point system has the problem of under flow _____ that limits the representation of very _____numbers.

8. The computation overhead for the fixed-point system is much _____ than that for the floating-point number system.

9. If x_1 is in $D(B, i, f)$ and x_2 is in $D(B, i, f)$, then $x_1 + x_2$ is in _____.

10. If x_1 is in $D(B, i_1, f_1)$ and x_2 is in $D(B, i_2, f_2)$ where $i_1 \neq i_2$ and $f_1 \neq f_2$ then it is _____ to add x_1 and x_2.

11. If x_1 is in $D(B, i, f)$ and x_2 is in $D(B, i, f)$, then $x_1 - x_2$ is in _____.

12. If x_1 is in $D(B, i_1, f_1)$ and x_2 is in $D(B, i_2, f_2)$ where $i_1 \neq i_2$ and $f_1 \neq f_2$ then it is _____ to subtract x_1 and x_2.

13. Multiplication can be performed on two numbers of _____ numeric formats but with the same base.

14. If x_1 is in $D(B, i_1, f_1)$ and x_2 is in $D(B, i_2, f_2)$, then product will be in _____.

DESCRIPTIVE QUESTIONS

1. Discuss how the fixed-point number system is represented.

2. For a 16-bit processor, what does the following mean?

$D(2, 8, 8)$, D(2, 7, 8), $D(2, 0, 16)$ and $D(2, 0, 15)$.

3. Write short notes on: (a) relative error, (b) relative density.

4. Discuss floating-point numeric format.

5. How does the floating-point representation differ from the fixed-point representation?

6. What is normalized floating-point?

7. How is zero represented in floating-point number system?

8. What is relative error and relative density with respect to floating-point number system?

9. Compare the relative errors and relative densities of fixed- and floating-point number systems.

10. What is underflow hole in floating-point number system?

11. How can the underflow hole problem in a floating-point number system be solved? Explain with an example.

12. For a 16-bit processor, what does the following mean?

$D(2, 12, 0, 7)$, $D(2, 12, -3, 3)$, $D(2, 14, 0, 3)$ and $D(2, 13, 0, 3)$.

13. What are the three primary arithmetic operators?

14. Discuss compatibility issues in the fixed-point numeric format of the numbers for addition, subtraction and multiplication operations.

15. If x_1 is in $D(B, i_1, f_1)$ and x_2 is in $D(B, i_2, f_2)$, then how to extract the product in $D(B, i_1, f_1)$?

16. If x_1 is in $D(B, i_1, f_1)$ and x_2 is in $D(B, i_2, f_2)$, then how to extract the product in $D(B, i_2, f_2)$?

17. What is normalization in fixed-point number representation?

18. What is scaling?

19. Discuss the normalization and scaling for converting an analog value to the digital number.

20. For a 16-bit system, if the variables in the application are ensured to have a value less than unity under all conditions, then all values including the result will be within the range of $D(2, 1, 14)$. Extend this result to an n-bit system.

21. Explain the multiplication algorithm.

22. What are the advantages of iterative algorithms?

23. Discuss the computation of the reciprocal and square roots using the closed-loop control method.

24. Write short notes on: (a) PI controller implementation, (b) sine and cosine implementation.

25. Write short notes on: (a) PWM for two-level two-arm bridge inverter, (b) space vector PWM.

26. Write short notes on: (a) circular boundary limit, (b) hexagonal boundary limit and (c) extending switching period.

PROBLEMS

1. Consider a 16-bit fixed-point system $D(2, 0, 15)$. What is its range? What is the relative error of the following equivalent decimal numbers: (a) 0.98536, (b) −0.234577, (c) 0.8379 and (d) 0.0011378.

2. Consider a 16-bit normalized floating-point system $D(2, 12, -3, 3)$. What is its range? What is the relative error of the following equivalent decimal numbers: (a) 0.9853637, (b) −0.23457733, (c) 0.8379 and (d) 0.0011378.

3. Perform multiplication for the following:
- **a.** $x_1 = +913.74$ in $D(10, 3, 2)$ and $x_2 = +908.14$ in $D(10, 3, 2)$
- **b.** $x_1 = +1.1354$ in $D(10, 1, 4)$ and $x_2 = +8.34$ in D(10, 1, 2)
- **c.** $x_1 = (+6.76846)_{10}$ in $D(2, 3, 12)$ and $x_2 = (+134.78)_{10}$ in $D(2, 8, 8)$
- **d.** $x_1 = (0.5)_{10}$ in $D(2, 0, 16)$ and $x_2 = (100)_{10}$ in $D(2, 8, 8)$
- **e.** $x_1 = (+1.54)_{10}$ in $D(2, 1, 14)$ and $x_2 = (-113.54)_{10}$ in $D(2, 7, 8)$

4. Derive the generic iteration algorithm for any given function $F(x, y) = 0$.

5. For a 16-bit binary number system, compute the reciprocal for the following numbers:
- **a.** $x = (5)_{10}$ in $D(2, 3, 12)$
- **b.** $x = (26)_{10}$ in $D(2, 6, 10)$
- **c.** $x = (4.5)_{10}$ in $D(2, 3, 12)$
- **d.** $x = (0.333)_{10}$ in $D(2, 1, 14)$
- **e.** $x = (10)_{10}$ in $D(2, 4, 12)$

6. For a 16-bit binary number system, compute the square roots for the following numbers:
- **a.** $x = (5)_{10}$ in $D(2, 3, 12)$

b. $x = (26)_{10}$ in $D(2, 6, 10)$
c. $x = (4.5)_{10}$ in $D(2, 3, 12)$
d. $x = (0.333)_{10}$ in $D(2, 1, 14)$
e. $x = (10)_{10}$ in $D(2, 4, 12)$

7. For a 16-bit binary number system, compute the reciprocal of the square roots for the following numbers:
a. $x = (5)_{10}$ in $D(2, 3, 12)$
b. $x = (26)_{10}$ in $D(2, 6, 10)$
c. $x = (4.5)_{10}$ in $D(2, 3, 12)$
d. $x = (0.333)_{10}$ in $D(2, 1, 14)$
e. $x = (10)_{10}$ in $D(2, 4, 12)$

8. For a 16-bit binary number system, compute the square root using the reciprocal of the square root iteration algorithm for the following numbers:
a. $x = (5)_{10}$ in $D(2, 3, 12)$
b. $x = (26)_{10}$ in $D(2, 6, 10)$
c. $x = (4.5)_{10}$in $D(2, 3, 12)$
d. $x = (0.333)_{10}$ in $D(2, 1, 14)$
e. $x = (10)_{10}$ in D(2, 4, 12)

9. For a 16-bit binary number system, compute e^x for the following numbers:
a. $x = (5)_{10}$ in $D(2, 3, 12)$
b. $x = (26)_{10}$ in $D(2, 6, 10)$
c. $x = (4.5)_{10}$ in $D(2, 3, 12)$
d. $x = (0.333)_{10}$ in $D(2, 1, 14)$
e. $x = (10)_{10}$ in $D(2, 4, 12)$

10. For a 16-bit binary number system, compute the hypotenuse by the open-loop method for the following values for the sides of the right-angled triangle:
a. $a = (5)_{10}$ in $D(2, 4, 12)$; $b = (7)_{10}$ in $D(2, 4, 12)$
b. $a = (26)_{10}$ in $D(2, 6, 10)$; $b = (13)_{10}$ in $D(2, 6, 10)$
c. $a = (4.5)_{10}$ in $D(2, 4, 12)$; $b = (7.5)_{10}$ in D(2, 4, 12)
d. $a = (0.333)_{10}$ in $D(2, 1, 14)$; $b = (0.895)_{10}$ in $D(2, 1, 14)$
e. $a = (10)_{10}$ in $D(2, 4, 12)$; $b = (5)10$ in $D(2, 4, 12)$

11. For a 16-bit binary number system, compute e^x for the following numbers:
a. $x = (5)_{10}$ in $D(2, 3, 12)$
b. $x = (26)_{10}$ in $D(2, 6, 10)$
c. $x = (4.5)_{10}$ in $D(2, 3, 12)$
d. $x = (0.333)_{10}$ in $D(2, 1, 14)$
e. $x = (10)_{10}$ in $D(2, 4, 12)$

12. For a 16-bit binary number system, compute $\ln(x)$ for the following numbers:
a. $x = (5)_{10}$ in $D(2, 3, 12)$
b. $x = (26)_{10}$ in $D(2, 6, 10)$
c. $x = (4.5)_{10}$ in $D(2, 3, 12)$
d. $x = (0.333)_{10}$ in $D(2, 1, 14)$
e. $x = (10)_{10}$ in $D(2, 4, 12)$

ANSWERS

Fill in the Blanks

1. range; density
2. overflow
3. resolution
4. uniform
5. base
6. zero
7. hole; small
8. lower
9. $D(B, i, f)$
10. not possible
11. $D(B, i, f)$
12. not possible
13. different
14. $D(B, i_1 + i_2, f_1 + f_2)$

CHAPTER 13

Thermal Aspects

Learning Objectives

After reading this chapter, you will be able to:

- understand the parameters used for thermal analysis.
- design heat sinks for a given application.

The power semiconductor devices dissipate power by virtue of the finite rise and fall times associated with the device switching. In Chapter 1 the method of calculating the power dissipation in a power semiconductor device is discussed. As a consequence, the junction temperature increases causing thermal stresses in the device. It is necessary to ensure that under no condition the junction temperature should exceed the maximum-rated junction temperature which is 150°C for most devices. This chapter discusses the issues involved in removing the heat generated within the semiconductor devices and the selection of heat sinks.

To obtain an insight into the thermal analysis of the power electronic systems, the heat flow mechanism must be understood. In general, there are four types of heat transfer mechanisms that are operative at any given time on a hot body. The degree to which a particular heat transfer mechanism dominates depends upon the specific application and the physical system. There are four types of heat transfer mechanisms which are as follows:

1. **Conduction:** Conduction is a process of heat transfer wherein the heat flows through solids from a higher temperature zone to a lower temperature zone.
2. **Convection:** Convection is the process of heat transfer by means of fluids.
3. **Radiation:** Radiation is a process of heat transfer, where the heat flows out from the hot body without the need for a heat transport medium.
4. **Mass Transport:** Mass transport is also a process of heat transfer wherein the heat is transferred by means of fluids. In a way it is similar to convection. However, convection is an uncontrolled process whereas mass transport is a controlled process where the discharge rate of the fluid can be controlled resulting in controlled heat transfer from one region to another.

It should be noted that all the heat transfer processes mentioned above are dependent on the properties of materials, fluids and the immediate environment.

13.1 Modeling Heat Transfer

To analyze the thermal aspects, one must first model the system from the thermal point of view. A system from the thermal point of view consists of a heat source, the heat transfer medium and the heat sink. In the case of the power electronics systems, the power semiconductor devices are not ideal switches. As discussed in Chapter 1, they dissipate power both during the ON-state as conduction losses and during the switching transition as switching losses. This causes the junction temperature of the power semiconductor devices to rise. The semiconductor junction of the power devices is considered as the heat source in most power electronic systems. The heat is transferred through the semiconductor material and the case of the device to the external ambient. The external ambient is the heat sink.

A generic thermal system is shown in Figure 13.1. It consists of a hot body, H at temperature T_1 at the center of a spherical enclosure. The spherical space within the enclosure is filled with a fluid such as air. The temperature at the inner surface of the enclosure is T_{si} and the temperature at the outer surface of the enclosure is T_{so}. At the far neighborhood of the spherical enclosure, the ambient temperature is T_A. The heat from the hot body flows outwards towards the spherical enclosure by radiation and convection. The heat flows through the solid enclosure wall by the conduction mechanism. External to the enclosure, the heat further flows into the atmosphere again by convection and radiation. It should be noted that as heat always flows from the higher temperature zone to the lower temperature zone, $T_1 > T_{si} > T_{so} > T_A$. The heat flow can be modeled as shown in Figure 13.2.

In the steady-state heat transfer schematic diagram shown in Figure 13.2 wherein the system is in thermal equilibrium, T_1 is the temperature of the hot body, T_{si} is the temperature at the inner surface of the spherical enclosure, T_{so} is the temperature at the outer surface of the enclosure and T_A is the ambient temperature. Also $R_{\theta v1}$ is the thermal resistance to heat flow between the hot body and the inner surface of the enclosure by the *convection* mechanism, $R_{\theta r1}$ is the thermal resistance to heat flow between the hot body and

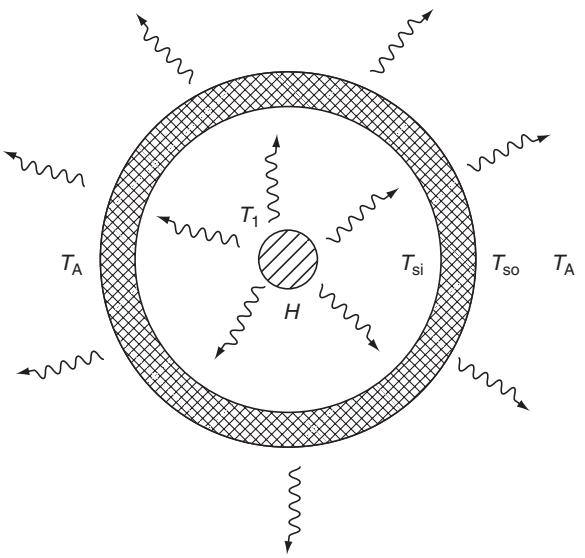


Figure 13.1 | Heat flow from hot body to ambient.

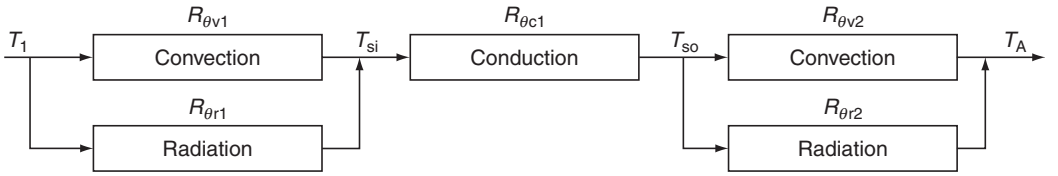


Figure 13.2 | Heat transfer model.

the inner surface of the enclosure by the *radiation* mechanism, $R_{\theta c1}$ is the thermal resistance to the heat flow between the inner and the outer surfaces of the enclosure by the *conduction* mechanism, $R_{\theta v2}$ is the thermal resistance to heat flow between the outer surface of the enclosure and the ambient by the *convection* mechanism and $R_{\theta r2}$ is the thermal resistance to heat flow between the outer surface of the enclosure and the ambient by the *radiation* mechanism. As the convection and radiation occur in parallel within and outside the enclosure, the corresponding thermal resistances are shown in parallel.

With reference to the heat flow indicated in Figure 13.2, the temperature difference is considered equivalent to the potential difference of electrical circuit and the rate of heat flow or power flow is considered equivalent to the current flow. It should be noted that the thermal domain is different from the other domains in that there is no concept of dissipation in this domain. The thermal resistance is not a dissipative element but only a notional element that indicates the conductivity of heat through the medium. There is no heat loss in the thermal resistance, but only a temperature difference. Higher the thermal resistance, lesser is the rate of heat flow in the medium and vice-versa. Thus, one can write

$$\Delta T = PR_\theta \tag{13.1}$$

where ΔT is the temperature difference in °K; P the heat power flow in W; R_θ the thermal resistance in °K/W. Equation (13.1) is sometimes referred to as the Ohm's Law for thermal domain. As an example, referring to Figure 13.2, the heat power flow through the convection mechanism from the hot body to the inside of the enclosure is given as

$$P_{\theta v1} = \frac{T_1 - T_{si}}{R_{\theta v1}} \tag{13.2}$$

Referring to Figure 13.3, A is the cross-sectional area of the medium that is orthogonal to the heat flow. It is usual to represent the rate of heat flow in terms of normalized power flow and is represented as

$$\frac{P}{A} = q \tag{13.3}$$

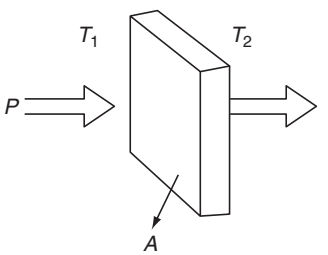


Figure 13.3 | Power flow and cross-section area.

where q is the rate of heat flow per unit area or specific heat flow rate and expressed in W/m^2. Applying Eq. (13.3) to the system schematic of Figure 13.2, one obtains

$$\frac{P_{\theta v1}}{A} = \frac{\Delta T}{AR_{\theta v1}} = \frac{\Delta T}{r_{\theta v1}} = q_{v1}$$

where $r_{\theta v1} = R_{\theta v1}A$ is called the *thermal resistivity* and is expressed in °Km^2/W. In general, one can write

$$q = \frac{\Delta T}{r_\theta} = h\Delta T \qquad (13.4)$$

where q is the specific heat flow rate in W/m^2; r_θ the thermal resistivity in °Km^2/W; $h = 1/r_\theta$ is called the *thermal coefficient* and expressed in W/°K/m^2. It should be observed that the material property of the medium is represented as R_θ, r_θ or h. In the literature, thermal resistance, thermal resistivity and thermal coefficient have been used in various circumstances; however, in general they represent the characteristic of the medium and can be used in any of the above forms and converted from one form to another using the above relationships. Like finding the power flow through the thermal resistance, $R_{\theta v1}$, the power flow through all the other branches can also be estimated in a manner similar to Eq. (13.2).

13.2 Heat Transfer by Conduction

Conduction is a heat transfer mechanism wherein heat energy flows from a hotter zone to a cooler zone in a solid. Heat transfer processes by conduction can be quantified in terms of the rate equation known as the Fourier's law. Consider a thin rectangular solid slab as shown in Figure 13.4.

Referring to Figure 13.4, q is the specific heat flow rate in the direction, T_1 is the temperature of the hotter side of the solid slab and T_2 is the temperature of the colder side of the slab. From Eq. (13.4),

$$q = h(T_1 - T_2) = h\Delta T \qquad (13.5)$$

where h is the thermal coefficient in W/°K/m^2. Let k be another constant called thermal conductivity which is defined in terms of the thermal coefficient and the thickness of the slab Δx. Thus,

$$k \equiv h\Delta x \quad [\text{W/°K/m}]$$

$$h = \frac{k}{\Delta x} \qquad (13.6)$$

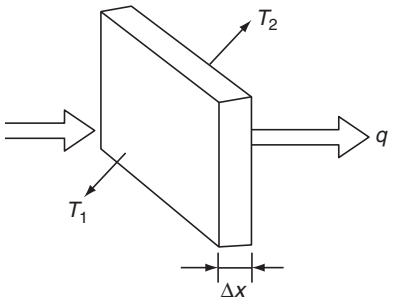


Figure 13.4 | Heat transfer through solids.

Substituting Eq. (13.6) in Eq. (13.5), one obtains

$$q = k\frac{\Delta T}{\Delta x} \tag{13.7}$$

Equation (13.7) is the heat flow rate equation. It implies a temperature gradient along a length of the solid slab in the direction of the heat flow. Thermal conductivity, k, is a transport property and is the characteristic of the solid slab material. If A is the cross-section area of the slab then the heat flow rate by conduction P_c is the product of the specific heat flow rate and the area as per Eq. (13.3). Thus,

$$P_c = kA\frac{\Delta T}{\Delta x} \tag{13.8}$$

The various thermal constants are summarized as follows:

1. Thermal resistance R_θ expressed in °K/W.
2. Thermal resistivity r_θ expressed in °Km^2/W.
3. Thermal coefficient h expressed in W/°K/m^2.
4. Thermal conductivity k expressed in W/°K/m.

The relationship of the various thermal constants in terms of thermal conductivity is as follows:

1. Thermal coefficient $h = k / \Delta x$ from Eq. (13.6).
2. Thermal resistivity

$$r_\theta = \frac{1}{h} = \frac{\Delta x}{k} \tag{13.9}$$

3. Thermal resistance

$$R_\theta = \frac{r_\theta}{A} = \frac{\Delta x}{kA} \tag{13.10}$$

Thermal conductivity is a material property. Table 13.1 gives the values of the thermal conductivity for some materials.

Table 13.1 | Thermal conductivities for some materials

Material	*k* value in W/°K/m
Silver	410
Copper	385
Aluminum	211
Steel	47.6
Glass	1.05
Heat-sink compound	0.4
Epoxy	0.2
Polyurethane	0.025
Polystyrene	0.035
Still air	0.026

Problem 13.1

Calculate thermal resistance to the flow of power through a solid slab wherein

(a) the slab is made up of aluminum having an area of 1 m^2 and thickness of 10 mm.
(b) the slab is made up of glass having an area of 1 m^2 and thickness of 5 mm.

Solution

The expression for the thermal resistance in terms of thermal conductivity is given as

$$R_\theta = \frac{\Delta x}{kA}$$

(a) $\Delta x = 10 \text{ mm} = 10 \times 10^{-3} \text{ m}$; $k = 211$; $A = 1 \text{ m}^2$

$$R_\theta = \frac{10 \times 10^{-3}}{211 \times 1} = 0.000004739 \text{ °K/W}$$

(b) $\Delta x = 5 \text{ mm} = 5 \times 10^{-3} \text{ m}$; $k = 1.05$; $A = 1 \text{ m}^2$

$$R_\theta = \frac{5 \times 10^{-3}}{1.05 \times 1} = 0.004762 \text{ °K/W}$$

A typical example of heat transfer by conduction is the case of heat dissipation in semiconductor devices such as junction transistors (BJT and MOSFET). Any device that has a voltage across it and current through it will dissipate power. This power is equal to the product of the voltage and the current. This power is dissipated as heat. In particular, a power semiconductor device that is switching will dissipate heat due to the ON-state loss and the switching loss as discussed in Chapter 1.

The maximum operating temperature of a semiconductor junction is usually 150°C. Exceeding this junction temperature limit will destroy the device. An important part of the design process is to ensure that the device junction temperature is within a certain specified limit. To safeguard this, *heat sinks* are used with semiconductor devices such that the thermal resistance is lowered to increase the rate at which the dissipated heat energy is drawn away from the junction to the external ambient. From Eq. (13.10), the thermal resistance R_θ is seen to depend on the thickness of the heat-sink material, thermal conductivity of the material and the area of the heat-sink material.

Once the heat-sink material is selected, the thermal conductivity being a property of the material is therefore determined for the material. The thickness and the surface area of the heat sink can be varied by the designer to achieve the thermal resistance for a particular application. Commercially, heat sinks with varied cross-sections are available as aluminum extrusions. Once the heat-sink cross-section for an application is decided, then the designer needs to only vary the length of the heat sink to achieve the required thermal resistance.

The heat generated in the junction should be removed in such a manner that at thermal equilibrium the junction temperature is well below the rated junction temperature. The electrical equivalent for the flow of heat from the junction to the ambient is depicted in Figure 13.5. In the steady-state equivalent circuit shown in Figure 13.5, $R_{\theta_{jc}}$ is the effective thermal resistance between the junction and the outer case of the device. This value is available in the datasheet of the specific device that is being used in the circuit. $R_{\theta cs}$ is the thermal resistance between the case and the heat sink and is dependent on the manner in which the

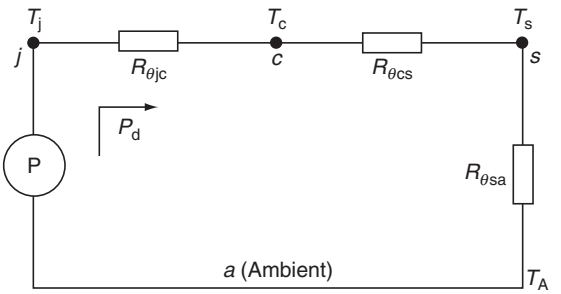


Figure 13.5 | Thermal equivalent circuit of a power semiconductor device mounted on a heat sink.

device is mounted on the heat sink. It varies between 0.1°K/W to 1°K/W depending on the type of mounting (such as non-isolated mounting, isolated mica washer mounting and skill of mounting). $R_{\theta sa}$ is the thermal resistance between the sink and the ambient. $R_{\theta sa}$ is to be calculated and a heat sink having a thermal resistance that is lower than the calculated $R_{\theta sa}$ should be selected.

A power semiconductor device mounted on a heat sink is shown in Figure 13.6. From Eq. (13.1), one obtains

$$T_j - T_A = P_d (R_{\theta jc} + R_{\theta cs} + R_{\theta sa}) \tag{13.11}$$

The values T_j, $R_{\theta jc}$ are known from datasheets. The heat sink should be selected such that even for the maximum ambient temperature, the power semiconductor switch's junction temperature is within safe limits. T_A should be taken at 50°C (worst case) as the ambient in the immediate neighborhood of the device's heat sink. P_d is the estimated dissipation in the power semiconductor switch. $R_{\theta cs}$ is between 0.1°K/W and 1°K/W. Thus $R_{\theta sa}$, that is, the thermal resistance between sink and ambient can be computed from Eq. (13.11) and an appropriate heat sink whose thermal resistance is less than the calculated value can be chosen. The concept of transient thermal impedance is important when dealing with pulsed overload currents. Mounting of the device on the heat sink and orientation of heat sink are to be considered carefully to increase the thermal efficiency and thereby improving the reliability of the device.

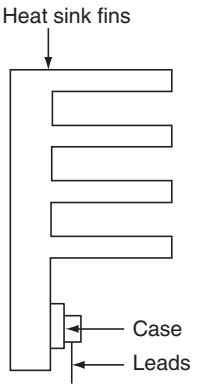


Figure 13.6 | A power semiconductor device mounted on a heat sink.

$$R_{\theta jc} + R_{\theta cs} + R_{\theta sa} = \frac{T_{jmax} - T_A}{P_{d\text{-}max}}$$

$$R_{\theta sa} = \frac{T_{jmax} - T_A}{P_{d\text{-}max}} - R_{\theta jc} - R_{\theta cs} \tag{13.12}$$

where T_{jmax} is the maximum allowable junction temperature which is usually assumed less than three-fourth of the maximum-rated junction temperature. Equation (13.12) is the design equation for selecting a heat sink considering only the conduction mechanism of heat transfer.

13.3 Heat Transfer by Convection

Convection is a mode of heat transfer between a solid and a fluid when there is a temperature difference. The convection heat transfer mode is composed of the following two mechanisms; (a) heat transfer due to random molecular motion (diffusion) and (b) heat transfer by the bulk motion of the fluid. Consider a heated surface as shown in Figure 13.7. The fluid above the heated surface is air in this case.

The fluid region close to the surface is at a temperature T_s and the temperature of the fluid at a farther distance from the surface is T_F. The higher temperature of the fluid closer to the surface makes the fluid molecules more agitated and less dense. As a result they rise up and the colder, denser, fluid portion reaches the surface which in turn will get heated up through interaction between the fluid and the heated surface. This results in the development of a region in the fluid through which the fluid velocity varies in the neighborhood of the surface.

The convection heat transfer mode is sustained both by diffusion and by the bulk motion of the fluid within the boundary layer. The contribution due to diffusion dominates near the surface where the fluid velocity is low. The contribution of bulk fluid motion is due to the fact that the boundary layer grows as the flow progresses in a direction orthogonal to the surface or in a direction away from the heated surface. In effect, the heat that is conducted into this layer in the immediate neighborhood of the surface is swept away and is eventually transferred to the fluid outside the immediate neighborhood.

Convection heat transfer may be classified into the following types:

1. **Free Convection:** Here the flow is induced by the density gradient which arises from density differences caused by temperature variations in the fluid.
2. **Forced Convection:** Here the flow is caused by external means such as fan, pump or atmospheric winds.

From Eq. (13.8), the rate of heat flow for the system of Figure 13.7 is

$$P = kA\frac{\Delta T}{\Delta x} \tag{13.13}$$

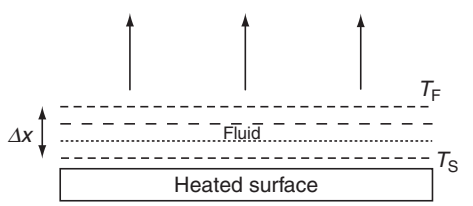


Figure 13.7 | Illustration of the convection mechanism.

where k is the thermal conductivity; A the cross-sectional area perpendicular to the heat flow; $\Delta T = T_S - T_F$ the temperature difference between the surface and outer fluid layer; Δx the thickness of the fluid in the immediate neighborhood of the heated surface. In Eq. (13.13), all the parameters are measurable except the fluid thickness Δx which is a quantity that has a measure of uncertainty in it. It varies with surroundings and external influences. An additional parameter X is included which is measurable and the fluid thickness is considered as a function of the measurable parameter X. Thus, Eq. (13.13) can be re-written as:

$$P = kA\left(\frac{X}{\Delta x}\right)\frac{\Delta T}{X} \tag{13.14}$$

where X is called the *characteristic dimension* of the heated solid surface. The ratio of the characteristic dimension to the fluid thickness is called the Nusselt number, a dimensionless quantity represented by $\mathcal{N}$. Eq. (13.14) is re-written as

$$P = kA\mathcal{N}\frac{\Delta T}{X} \tag{13.15}$$

Rearranging the terms of Eq. (13.15), one obtains the expression for heat flow rate/unit area and convective thermal resistance as

$$\frac{P}{A} = q = k\mathcal{N}\frac{\Delta T}{X} \tag{13.16}$$

$$R_{\theta v} = \frac{\Delta T}{P} = \frac{X}{k\mathcal{N}A} \tag{13.17}$$

From Eq. (13.9), the thermal resistivity is given as

$$r_{\theta v} = R_{\theta v}A = \frac{X}{k\mathcal{N}} \tag{13.18}$$

The convective thermal coefficient is the reciprocal of the thermal resistivity and therefore,

$$h_{\theta v} = \frac{1}{r_{\theta v}} = \frac{k\mathcal{N}}{X} \tag{13.19}$$

Substituting for convective thermal coefficient in the expression for heat flow rate of Eq. (13.16), one obtains

$$q = h_{\theta v}\Delta T$$

$$P = h_{\theta v}A\Delta T \tag{13.20}$$

Determination of Nusselt Number

In the determination of the convective thermal resistance or the thermal coefficient, the factor with a measure of uncertainty is the Nusselt number. The Nusselt number is a dimensionless quantity which is a function of the velocity of the fluid flow, the properties of the fluid medium and the geometry of the hot solid body which transfers the heat to the fluid. Further, the Nusselt number also depends on the type of convection, whether free convection or forced convection. In the case of free convection, one uses another dimensionless quantity called the Rayleigh number to compute the Nusselt number. In the case of forced convection, one uses the dimensionless quantity called the Reynolds number to compute the Nusselt number.

Rayleigh Number

This quantity is required for free convection. It is represented by $\mathcal{A}$ and is given by the following expression:

$$\mathcal{A} = \frac{g\beta X^3 \Delta T}{\delta \cdot \nu} \tag{13.21}$$

where g is the acceleration due to gravity = 9.81 m/s^2; β the coefficient of thermal expansion of the fluid medium; X the characteristic dimension of the hot solid surface; δ the thermal diffusivity; ν the kinematic viscosity of the fluid medium.

Reynolds Number

This quantity is required for forced convection. It is represented by $\mathcal{R}$ and is given by the following expression:

$$\mathcal{R} = \frac{uX}{\nu} \tag{13.22}$$

where u is the mean velocity of forced fluid flow; X the characteristic dimension of the hot solid surface; ν the kinematic viscosity of the fluid medium.

The Nusselt number is a function of both the Rayleigh number and Reynolds number and is given as

$$\mathcal{N} = f(\mathcal{A}, \mathcal{R}) \tag{13.23}$$

When the fluid motion is highly ordered and has a streamlined flow, then such fluid motion is called laminar flow. If the fluid motion is highly irregular and is characterized by velocity fluctuations, then such a fluid motion is called turbulent flow. The Rayleigh number and the Reynolds number have been used for classifying whether a fluid flow is laminar or turbulent. The laminar–turbulent distinction is dependent on the solid–fluid interface geometries and may vary for various geometries and positions.

Free Convection

The Nusselt number is best found by experimentation for an arbitrary shape of the hot solid. However, for standard shapes of the solids, the following curve-fit relationships can be used for preliminary designs. These should, however, be verified and fine tuned after experimentation as there is always a certain degree of uncertainty in estimating the flows and the nature of the flows, whether laminar or turbulent. For any convective-based thermal system, both the characteristic dimension X of the solid and the Nusselt number for the fluid are needed to evaluate the thermal coefficient as per Eq. (13.19).

Horizontal Flat Plate

Figure 13.8 illustrates horizontal flat plates of two different geometries, viz., a rectangular geometry with dimensions a and b as shown and another having a circular geometry with diameter d as indicated.

The thickness of the plate is considered small compared to the other indicated dimensions. The plate is placed horizontal and is the heat source. The fluid (air) heats up near the hot surface, becomes less dense and moves up as shown. The Nusselt numbers for the plates of geometry given in Figure 13.8 are computed as follows: The characteristic dimension X for the rectangular plate is given as

$$X = \frac{a+b}{2}$$

and for the circular plate

$$X = d$$

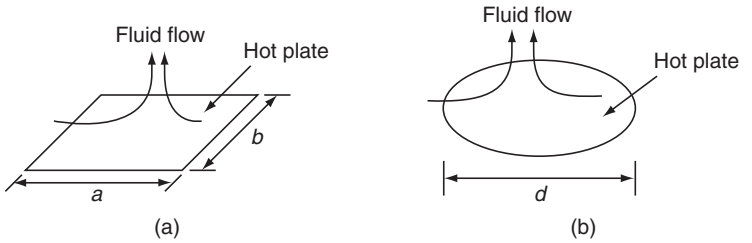


Figure 13.8 | Horizontal flat plate: (a) Rectangular plate; (b) circular plate.

For a laminar flow where $10^2 < \mathcal{A} < 10^5$, the Nusselt number is determined as

$$\mathcal{N} = 0.54\, \mathcal{A}^{0.25} \tag{13.24}$$

For a turbulent flow where $\mathcal{A} > 10^5$, the Nusselt number is determined as

$$\mathcal{N} = 0.14\, \mathcal{A}^{0.33} \tag{13.25}$$

Vertical Flat Plate

Figure 13.9 illustrates vertical flat plates of two different geometries, viz., a rectangular plate with dimensions *a* and *b* as shown and another having a circular geometry with diameter *d* as indicated.

The thickness of the plate is considered small compared to the other indicated dimensions. The plate is placed vertical and is the heat source. The fluid (air) heats up near the hot surface, becomes less dense and moves up as shown. The Nusselt numbers for the plates are computed as follows: The characteristic dimension *X* for a rectangular plate is given as

$$X = b$$

and for a circular plate

$$X = d$$

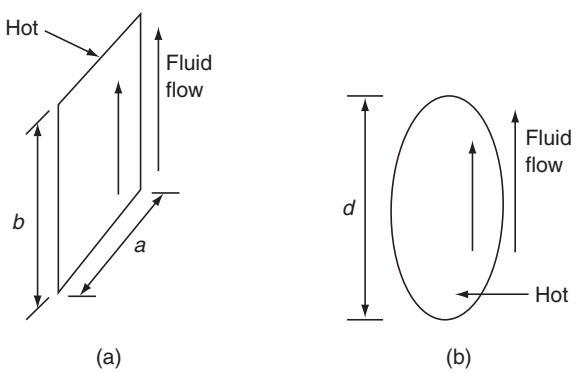


Figure 13.9 | Vertical flat plate: (a) Rectangular plate; (b) circular plate.

For a laminar flow where $10^4 < \mathcal{A} < 10^9$, the Nusselt number is determined as

$$\mathcal{N} = 0.56\, \mathcal{A}^{0.25} \tag{13.26}$$

For a turbulent flow where $10^9 < \mathcal{A} < 10^{12}$, the Nusselt number is determined as

$$\mathcal{N} = 0.20\, \mathcal{A}^{0.40} \tag{13.27}$$

From Eqs. (13.24) to (13.27), it can be observed that a flat plate placed vertically results in a higher Nusselt number and consequently the thermal resistance is lower. This implies that the vertical plate will convect the heat away from the solid surface better than from a horizontal surface.

Horizontal Cylindrical Solid

Figure 13.10 illustrates a horizontally placed cylindrical solid with diameter d and length l as shown.

The cylinder is the hot body and the heat source. The fluid (air) heats up near the hot surface, becomes less dense and moves up as shown. The Nusselt number for the horizontal cylindrical geometry given in Figure 13.10 is computed as follows: The characteristic dimension X is given as

$$X = d$$

For a laminar flow where $10^4 < \mathcal{A} < 10^9$, the Nusselt number is determined as

$$\mathcal{N} = 0.47\, \mathcal{A}^{0.25} \tag{13.28}$$

For a turbulent flow where $\mathcal{A} > 10^9$, the Nusselt number is determined as

$$\mathcal{N} = 0.1\, \mathcal{A}^{0.33} \tag{13.29}$$

Vertical Cylindrical Solid

Figure 13.11 illustrates a vertically placed cylindrical solid with diameter d and length l as shown. The cylinder is the hot body and the heat source. The fluid (air) heats up near the hot surface, becomes less dense and moves up as shown. The Nusselt number for the vertical cylindrical geometry given in Figure 13.11 is computed as follows: The characteristic dimension X is given as

$$X = l$$

For a laminar flow where $10^4 < \mathcal{A} < 10^9$, the Nusselt number is determined as

$$\mathcal{N} = 0.56\, \mathcal{A}^{0.25} \tag{13.30}$$

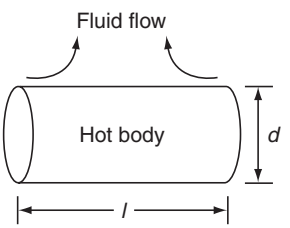


Figure 13.10 | Horizontally placed solid cylinder.

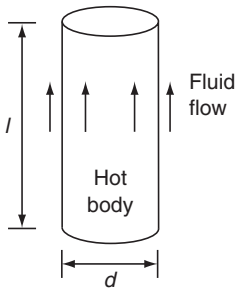


Figure 13.11 | Vertically placed solid cylinder.

For a turbulent flow where $10^9 < \mathcal{A} < 10^{12}$, the Nusselt number is determined as

$$\mathcal{N} = 0.20\ \mathcal{A}^{0.40} \tag{13.31}$$

From Eqs. (13.28) to (13.31), it can be observed that a solid cylinder placed vertically results in a higher Nusselt number and consequently lower thermal resistance as compared to a horizontally placed cylinder.

Parallel Flat Plates

Consider two plates placed parallel with respect to each other as shown in Figure 13.12(a). In Figure 13.12(b), the parallel plates are at an angle of θ with respect to the horizontal. The two parallel flat plates are separated by a distance d.

The plates have a difference in temperature ΔT with respect to each other as indicated. The characteristic dimension for the parallel plate system is

$$X = d$$

As the fluid flow within the parallel plates is not laminar, only turbulent flows are considered for this configuration. For slope angles, $\theta < 50°$ the Nusselt number for turbulent fluid flow ($\mathcal{A} > 10^5$) is determined as

$$\mathcal{N} = 0.062\ \mathcal{A}^{0.33} \tag{13.32}$$

Forced Convection

In forced convection, a controlled fluid flow is applied along the surface of the solid hot body to improve the convective heat transfer. A fan is usually employed to create the controlled fluid flow. The fluid flow is forced

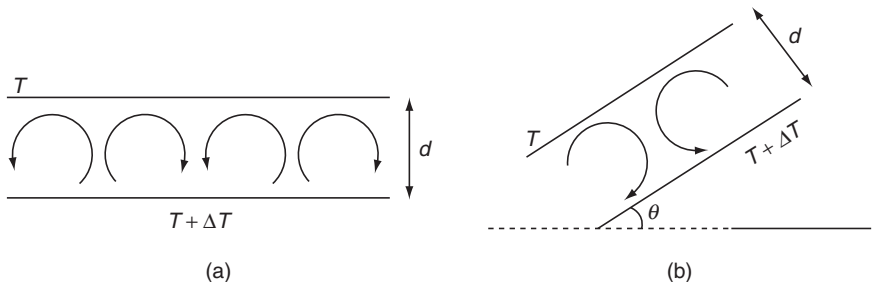


Figure 13.12 | Parallel plates: (a) Horizontal; (b) at a slope with respect to the horizontal.

along the surface of the solid hot body irrespective of the orientation of the hot surface. Therefore, orientation is not as significant in forced convection as is the case in free convection. In forced convection, the Reynolds number is used to compute the Nusselt number.

Flat Plate

Consider a flat plate of any geometry as shown in Figure 13.13. A fan is used to force a fluid flow over the hot surface which has a length a along the flow as shown.

The characteristic dimension of the flat plat is the length of the flat plate along the direction of the flow. This is given as

$$X = a$$

For a laminar flow where $\mathcal{R} < 5 \times 10^5$, the Nusselt number is determined as

$$\mathcal{N} = 0.664 \times \mathcal{R}^{0.5} \times (v/\delta)^{0.33} \tag{13.33}$$

For a turbulent flow where $\mathcal{R} > 5 \times 10^5$, the Nusselt number is determined as

$$\mathcal{N} = 0.37 \times \mathcal{R}^{0.8} \times (v/\delta)^{0.33} \tag{13.34}$$

Solid Cylinder

Consider a cylindrical solid as shown in Figure 13.14. A fan is used to force a fluid flow over the hot surface of the cylinder as shown.

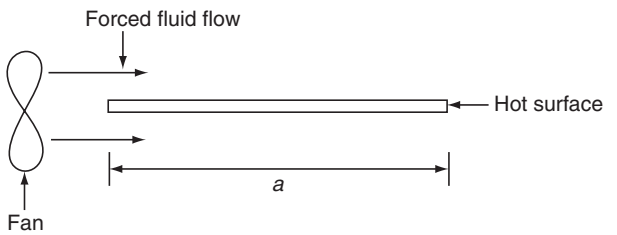


Figure 13.13 | Forced convection over flat plate.

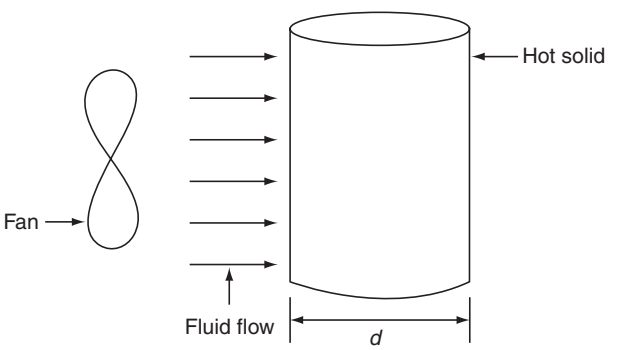


Figure 13.14 | Forced convection over cylinder.

The characteristic dimension of the solid cylinder is given as

$$X = d$$

For a laminar flow where $0.1 < \mathcal{R} < 1000$, the Nusselt number is determined as

$$\mathcal{N} = (0.35 + 0.56\ \mathcal{R}^{0.52}) \times (v/\delta)^{0.3} \tag{13.35}$$

For a turbulent flow where $1000 < \mathcal{R} < 5 \times 10^5$, the Nusselt number is determined as

$$\mathcal{N} = 0.26\ \mathcal{R}^{0.6} \times (v/\delta)^{0.3} \tag{13.36}$$

Problem 13.2

Consider a power BJT mounted on a flat square aluminum plate measuring 100 mm on each side. The flat plate is part of the enclosure of the product as shown in Figure 13.15. Let the temperature within the enclosure *be 45°C and the external ambient at a distance from the flat plate is also at 45°C. The power BJT is operating in the linear region and dissipates power. As a consequence, the temperature of the square aluminum plate is 70°C. Calculate the heat flow in watts from the plate to the external ambient by convection through the air.*

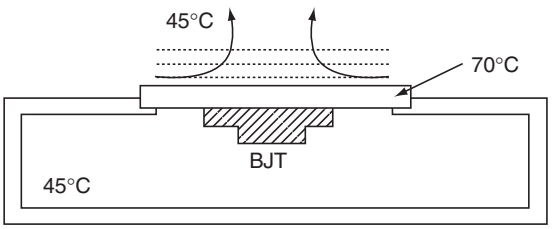


Figure 13.15 | Problem 13.2.

Solution

From Eqs. (13.19) and (13.20), the power flow P is given by

$$P = A\frac{k\mathcal{N}}{X}\Delta T$$

where A is the area of the surface perpendicular to the heat flow = 0.01 m^2; $\mathcal{N}$ the Nusselt number; k the thermal conductivity of the fluid (air) = 0.028 W/°K/m; $X = (a + b)/2 = 0.1$ m for the horizontal flat plate; ΔT the temperature difference between the plate and the external ambient = 70 − 45 = 25°K. Considering free convection, one needs to first compute the Rayleigh number ($\mathcal{A}$) in order to evaluate the Nusselt number. $\mathcal{A}$ is given by Eq. (13.21) as

$$\mathcal{A} = \frac{g\beta X^3 \Delta T}{\delta \cdot \upsilon}$$

where g is the acceleration due to gravity = 9.81 m/s^2; β the coefficient of thermal expansion of the fluid (air) = 1/330°K; X the characteristic dimension for flat plat = 0.1 m; $\Delta T = (273 + 70)°K - (273 + 45)°K = 25°K$; δ the thermal diffusivity of the fluid (air) = 2.6×10^{-5} m^2/s (from scientific tables); v the kinematic viscosity of the fluid (air) = 1.8×10^{-5} m^2/s (from scientific tables). Using these values,

$$\mathcal{A} = \frac{(9.81\ \text{m/s}^2)(1/330^\circ\text{K})(0.1\ \text{m})^3 25^\circ\text{K}}{(2.6\times10^{-5}\ \text{m}^2/\text{s})(1.8\times10^{-5}\ \text{m}^2/\text{s})}$$

$$\mathcal{A} = 1.588 \times 10^6$$

Since $\mathcal{A} > 10^5$, the flow is turbulent. The Nusselt number is given by Eq. (13.25) as

$$\mathcal{N} = 0.14\ \mathcal{A}^{0.33}$$

$$\mathcal{N} = 0.14 \times (1.588 \times 10^6)^{0.33} = 15.57$$

Now, substituting the value of Nusselt number in the expression for power flow (P) one obtains the rate of heat flow or the power flow as

$$P = (0.01\ \text{m}^2)(0.028\ \text{W/m}^\circ\text{K})\left(\frac{15.57}{0.1\ \text{m}}\right)(25^\circ\text{K}) = 1.09\ \text{W}$$

Problem 13.3

Consider a MOSFET mounted underneath a solid aluminum cylindrical heat sink as shown in Figure 13.16. The cylinder has a diameter of 22 cm and a height of 11 cm. The ambient temperature is 20°C and the space within the enclosure is also assumed to be at the ambient temperature. The cylinder is at a temperature of 100°C due to the dissipation from the heat source, that is, the MOSFET mounted underneath the cylindrical heat sink. Calculate the energy required to maintain heat-sink temperature at 100°C under the following conditions:

(a) Free convection with almost still air environment for the system as shown in Figure 13.16(a).
(b) Forced convection with a fan mounted that provides an air flow of 3 m/s for the system as shown in Figure 13.16(b).

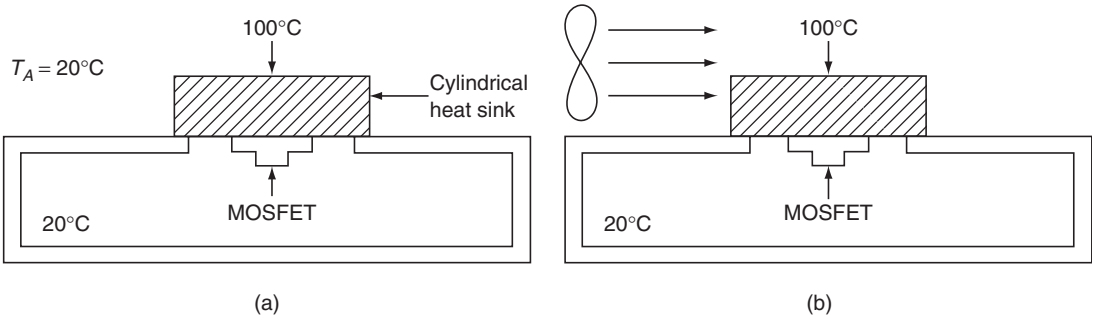


Figure 13.16 | Problem 13.3.

Solution

Case (a): This is a free convection problem. Therefore, the Rayleigh number needs to be evaluated to calculate Nusselt number. The energy from the MOSFET is transferred to the ambient due to convection in two directions: (a) through the top portion of the cylinder and (b) through the sides of the cylinder.

It is assumed that the body of the solid cylinder has negligible thermal resistance to conduction from MOSFET to the cylinder mass as compared to the convection thermal resistance from the cylinder to the surrounding air. The cylindrical solid may be resolved into standard surfaces to solve the problem. The top portion of the cylindrical heat sink is considered as a circular horizontal flat plate. The sides are considered as a cylindrical surface.

Top portion of the heat sink: The top portion is considered as a circular flat plate. The Rayleigh number is calculated from

$$\mathcal{A}_{\text{top}} = \frac{g\beta X^3 \Delta T}{\delta \cdot \upsilon}$$

where g is the acceleration due to gravity = 9.81 m/s^2; β the coefficient of thermal expansion of the fluid (air) = 1/330°K; X the characteristic dimension of the circular plate which is the diameter of the plate = 0.22 m; $\Delta T = (273 + 100)°\text{K} - (273 + 20)°\text{K} = 80°\text{K}$; δ the thermal diffusivity of the fluid (air) = 2.6×10^{-5} m^2/s (from scientific tables); υ the kinematic viscosity of the fluid (air) = 1.8×10^{-5} m^2/s (from scientific tables). Using these values,

$$\mathcal{A}_{\text{top}} = \frac{(9.81\ \text{m/s}^2)(1/330^\circ\text{K})(0.22\ \text{m})^3(80^\circ\text{K})}{(2.6\times10^{-5}\ \text{m}^2/\text{s})(1.8\times10^{-5}\ \text{m}^2/\text{s})}$$

$$\mathcal{A}_{\text{top}} = 5.41087 \times 10^7$$

Since $\mathcal{A}_{\text{top}} > 10^5$, the Nusselt number is calculated from Eq. (13.25) as

$$\mathcal{N} = 0.14\ \mathcal{A}^{0.33} = 0.14\ (5.41087 \times 10^7)^{0.33} = 49.901$$

The expression for the rate of heat flow is given by

$$P = A\frac{\mathcal{N}}{X}k\Delta T$$

where A is the surface area of the circular plate given by $A = (\pi/4)d^2$, d being the diameter of the plate. Substituting the appropriate values into the above power flow equation, one obtains

$$P_{\text{top}} = \left(\frac{\pi}{4}\right)(0.22\ \text{m})^2\left(\frac{49.901}{0.22\ \text{m}}\right)(0.028\ \text{W/m°K})(80°\text{K}) = 19.3139\ \text{W}$$

Side portion of the heat sink: The side portion of the heat sink is considered as a vertical cylinder where the characteristic dimension X is the height of the cylinder. Substituting the appropriate values into the equation for the Rayleigh number, one obtains

$$\mathcal{A}_{\text{side}} = \frac{(9.81\ \text{m/s}^2)(1/330^\circ\text{K})(0.11\ \text{m})^3(80^\circ\text{K})}{(2.6\times10^{-5}\ \text{m}^2/\text{s})(1.8\times10^{-5}\ \text{m}^2/\text{s})} = 6.7636\times10^6$$

Since $10^4 < \mathcal{A} < 10^9$, the Nusselt number is given by Eq. (13.30) as

$$\mathcal{N} = 0.56\ \mathcal{A}^{0.25} = 0.56 \times (6.7636 \times 10^6)^{0.25} = 28.5583$$

The surface area of the side of the cylinder is given by $A = \pi dh$, where d is the diameter of the cylinder and h is the height. Substituting the appropriate values into the power flow equation, one obtains

$$P_{\text{side}} = (\pi \times 0.22\ \text{m} \times 0.11\ \text{m})\left(\frac{28.5583}{0.11\ \text{m}}\right)(0.028\ \text{W/m°K})(80°\text{K}) = 44.2133\ \text{W}$$

The total heat flow is the sum of the above two. That is

$$P_{\text{total}} = P_{\text{top}} + P_{\text{side}} = 19.3139 \text{ W} + 44.2133 \text{ W} = 63.5272 \text{ W}$$

Hence, the energy required to maintain the heat sink at 100°C for 1 h is

$$63.5272 \text{ W} \times 1 \text{ h} = 63.5272 \text{ Wh}$$

Case (b): This is a forced convection problem where the air flows at the velocity of 3 m/s over the cylindrical heat sink. Reynolds number needs to be evaluated to find the Nusselt number. The expression for Reynolds number is given by

$$\mathcal{R} = \frac{uX}{\upsilon}$$

where u is the average velocity in m/s = 3 m/s; X the characteristic dimension in m and v the kinematic viscosity of the fluid in m^2/s. As in the case of free convection, the loss of energy is from the top and sides. The loss of energy in each of these directions can be found as follows:

Top portion of the heat sink: The top portion is considered as a circular flat plate with air flow over the plate. The characteristic dimension X of the circular plate is its diameter. The Reynolds number for this plate is

$$\mathcal{R} = \frac{(3 \text{ m/s})(0.22 \text{ m})}{(1.8 \times 10^{-5} \text{ m}^2/\text{s})} = 0.3667 \times 10^5$$

Since $\mathcal{R} < 5 \times 10^5$, the flow is laminar. The Nusselt number is found from the following expression using Eq. (13.33),

$$\mathcal{N} = 0.664 \times \mathcal{R}^{0.5} \times (\nu/\delta)^{0.33}$$

where δ is the thermal diffusivity of the fluid.
Substituting the appropriate values, the Nusselt number is

$$\mathcal{N} = (0.664)(0.3667 \times 10^5)^{0.5} \left(\frac{1.8 \times 10^{-5} \text{ m}^2/\text{s}}{2.6 \times 10^{-5} \text{ m}^2/\text{s}} \right)^{0.33} = 112.622$$

The Nusselt number is used to calculate the heat flow rate from the following expression:

$$P_{\text{top}} = A \frac{N}{X} k \Delta T$$

where A is the area of the circular plate = $\pi d^2 / 4$ m^2; k is the thermal conductivity of the fluid and ΔT is the temperature difference. Substituting the appropriate values, one obtains

$$P_{\text{top}} = \left(\frac{\pi}{4} \right) (0.22 \text{ m})^2 \left(\frac{112.622}{0.22 \text{ m}} \right) (0.028 \text{ W/m°K})(80 \text{ °K}) = 43.5897 \text{ W}$$

Side portion of the heat sink: The side portion constitutes a vertical cylinder with a forced air flow over it. The characteristic dimension X for this case is the diameter of the cylinder. Hence the Reynolds number in this case also would be same as that calculated for the top portion.

$$\mathcal{R} = 0.3667 \times 10^5 \text{ or } 3.667 \times 10^4$$

Since $1000 < \mathcal{R} < 5 \times 10^4$, the flow is considered turbulent. The Nusselt number is evaluated from the following expression as given by Eq. (13.36):

$$N = 0.26R^{0.6}\left(\upsilon / \delta\right)^{0.3} = (0.26)(3.667\times10^4)^{0.6}\left(\frac{1.8\times10^{-5}\ \text{m}^2/\text{s}}{2.6\times10^{-5}\ \text{m}^2/\text{s}}\right)^{0.3} = 127.541$$

and the heat flow rate is given as

$$P_{\text{side}} = (\pi \times 0.22\ \text{m} \times 0.11\ \text{m})\left(\frac{127.541}{0.22\ \text{m}}\right)(0.028\ \text{W/m°K})(80°\text{K}) = 98.728\ \text{W}$$

The total heat flow is the sum of the heat flow through the top and through the sides.

$$P_{\text{total}} = (43.5897 + 98.728)\ \text{W} = 142.3177\ \text{W}$$

Hence, the energy required to maintain the heat sink at 100°C for 1 h is

$$142.3177\ \text{W} \times 1\ \text{h} = 142.3177\ \text{Wh}$$

It can be observed that the forced convection handles greater power dissipation than the free convection for the same heat-sink dimensions and heat-sink temperature. It should be noted that in real situations even in the presence of the fan, both the free and forced convection are active simultaneously and the resultant total power flow is not simply an algebraic addition as indicated above. Further, it should be observed that in the calculation of the Reynolds number, the flow over the flat plate on the top is laminar whereas the flow on the sides of the cylinder is turbulent. The demarcation between the laminar and the turbulent flow along the cylinder height is not very distinct and therefore the calculated results obtained as above should be considered with a measure of uncertainty. The actual values of the Nusselt number or the power flow should be obtained only by experiment for greater accuracy. However, the above method will give a preliminary design that can be used to size the components of the products reasonably.

13.4 Heat Transfer by Radiation

A substance at a finite temperature will emit energy called thermal radiation. The heat energy is transported by electromagnetic waves. The transfer of heat energy by radiation does not require the presence of a material medium. In fact, radiation transfer occurs most efficiently in vacuum.

Consider a hot body at temperature T_1 as shown in Figure 13.17. Let T_2 be the temperature at some imaginary horizontal boundary parallel to the radiating surface such that $T_2 < T_1$. From Eq. (13.4), the rate of heat energy flow is given as

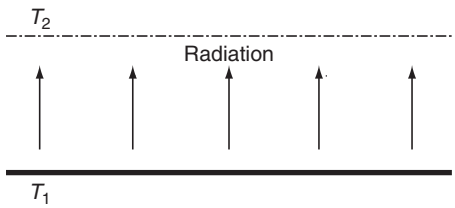


Figure 13.17 | Heat transfer by radiation.

Table 13.2 | Emittances for common radiation surfaces

Material type	Emittance
Anodized aluminum (black)	0.800
Polished aluminum	0.095
Rough surface aluminum	0.180
Rough surface iron	0.170
Tungsten at 1500°C	0.330
Enamel paint	0.85
Polished copper	0.07
Glass	0.940
Wood	0.900

$$P = h_r A \Delta T$$

where h_r is the radiation thermal coefficient; A the surface area perpendicular to the radiation direction; ΔT the temperature difference between the radiating body and the reference region.

The radiating thermal coefficient is given by the following expression:

$$h_r = 4\sigma\varepsilon_{eff}(1-\phi)\left(\frac{T_1+T_2}{2}\right)^3 \tag{13.37}$$

where σ is the Boltzmann's constant = 5.67×10^{-8} W/m^2/°K^4; ε_{eff} the effective emittance; ϕ the shielding factor. The shielding factor ϕ gives a measure of the shading of the radiative heat transfer from a hot body due to the presence of another body in the vicinity. For a single plate or two parallel plates, the shielding factor is zero.

The emittance is the energy emitted relative to that of a black body. Its value is less than unity. The emittance depends on the properties such as surface type and the color of the material. Table 13.2 gives the emittance of some of the commonly used materials. A material having an emittance of 1 is called a black body and is a perfect radiator. It also means that the material cannot retain any heat and the heat would be lost completely over a period. Such materials would be useless in applications where heat retention is required.

Problem 13.4

Consider two parallel plates each having area of 1 m^2. Let the first plate be made up of glass and kept at a temperature of 350°K. Let the second plate be made up of rough surfaced aluminum. Let the temperature of the second plate be at 300°K. Find the radiation heat flow from the first plate to the second plate.

Figure 13.18 shows the arrangement of the two plates.

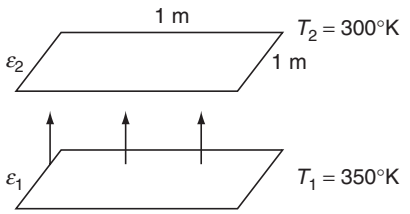


Figure 13.18 | Problem 13.4.

Solution

The bottom plate is a glass plate which is at 350°K and from Table 13.2 the emittance of glass is 0.940. The top plate is made up of rough surfaced aluminum having emittance of 0.18. The temperature of that plate is 300°K. The radiation heat flow is from plate 1 to plate 2. The radiation thermal coefficient of the path between plate 1 and plate 2 can be calculated from the following expression:

$$h_r = 4\sigma\varepsilon_{eff}(1-\phi)\left(\frac{T_1+T_2}{2}\right)^3 \text{ W/m}^2/°\text{K}$$

where σ is the Boltzmann's constant = 5.67×10^{-8} W/m^2/°K^4;
ϕ the shielding factor = 0 for parallel plates;

$$\varepsilon_{eff} = \frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2-\varepsilon_1\varepsilon_2}$$

for parallel plates. Using the above values, the thermal coefficient is given as

$$h_r = 4(5.67\times10^{-8}\ \text{W/m}^2/°\text{K}^4)\left(\frac{(0.94)(0.18)}{0.94+0.18-(0.94)(0.18)}\right)(1-0)\left(\frac{350°\text{K}+300°\text{K}}{2}\right)^3$$

$$h_r = 1.3855 \text{ W/m}^2/°\text{K}$$

The radiation heat flow rate can now be calculated from the following expression:

$$P = h_r A\Delta T = (1.3855\ \text{W/m}^2/°\text{K})(1\ \text{m}^2)(350°\text{K}-300°\text{K}) = 69.275\ \text{W}$$

This implies that 69.275 W of power is lost due to radiation from plate 1 to plate 2.

Reflectance, Absorptance and Transmittance

Whenever a wave of energy hits a surface, part of the energy gets reflected, part of the energy is absorbed by the surface and part of the energy gets transmitted depending on the type of the surface. The amount of energy that gets reflected depends on a factor called the *reflectance* of the surface denoted by ρ. The amount of energy that is absorbed depends on a factor called the *absorptance* of the surface denoted by α. The amount of energy that gets transmitted depends on a factor called *transmittance* of the surface denoted by τ. For any surface, the following relationship holds good:

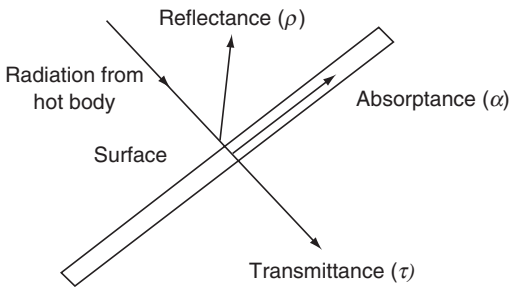


Figure 13.19 | Reflectance, absorptance and transmittance.

$$\rho + \alpha + \tau = 1 \tag{13.38}$$

The relationship in Eq. (13.38) is deduced from the conservation of energy principle and these constants depend on the type of surface. These three effects are illustrated in Figure 13.19 where radiated energy from a hot body strikes the surface.

As an example, for a glass surface, α (absorptance) is almost equal to 0; the transmittance τ is high and equal to 0.92; the reflectance ρ is low and equal to 0.08. The sum of these factors is equal to 1, satisfying the above relationship.

13.5 Heat Transfer by Mass Transport

Heat transfer by mass transport also involves the solid–fluid interaction. The solid is the heat source and the heat from it is transferred to the fluid which takes away the heat energy. The flow rate of the fluid can be used for controlling the heat energy transfer. This method of heat transfer is used in water-cooled heat sinks to achieve improved heat power removal.

Consider a fluid flow through a heated pipe as shown in Figure 13.20. The pipe is heated by means of a heat source which could be the power dissipated by a power semiconductor device or any dissipative device. Let T_p be the temperature of the pipe. As the pipe is at a higher temperature compared to the ambient, there will exist convection and radiation from the pipe surface to the external ambient. In addition to the power transferred to

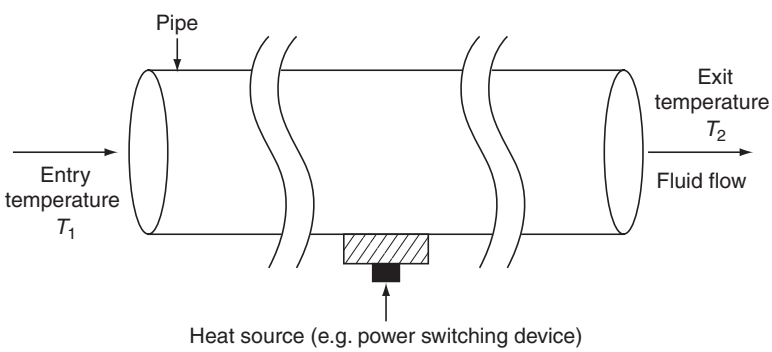


Figure 13.20 | Heat transfer by mass transport.

the external ambient through convection and radiation, power is transferred from the pipe to the flowing fluid also. This section discusses the amount of heat energy that is transferred from the pipe to the flowing fluid.

Let T_1 be the temperature of the fluid at the entry region of the pipe. Let T_2 be the temperature at the exit region such that $T_1 < T_2$. Let m be the mass of the fluid flowing through the pipe and $\mathrm{d}m/\mathrm{d}t$ the mass flow rate in kg/s. The heat flow rate due to the transfer of a mass m of the fluid from one place to another is called the heat transfer by mass transport. Let P_m be the power transferred, which is given as

$$P_m = \frac{\mathrm{d}m}{\mathrm{d}t} c_p (T_2 - T_1) \tag{13.39}$$

where c_p is called the specific heat of the fluid and is expressed in J/kg/°K which is a constant for a given fluid and $\mathrm{d}m/\mathrm{d}t$ is the mass flow rate. The thermal resistance of the mass transport process can be computed as follows: Let R_m be the equivalent thermal resistance to heat transfer by mass transport. Then from Eq. (13.1),

$$T_2 - T_1 = P_m R_m \tag{13.40}$$

Therefore

$$R_m = \frac{T_2 - T_1}{P_m} \tag{13.41}$$

Substituting the value of P_m from Eq. (13.39) into Eq. (13.41), one obtains

$$R_m = \frac{T_2 - T_1}{(\mathrm{d}m/\mathrm{d}t) c_p (T_2 - T_1)} = \frac{1}{(\mathrm{d}m/\mathrm{d}t) c_p} \tag{13.42}$$

It can be observed here that temperature is not the driving function for the heat transfer unlike conduction, free convection and radiation. The heat flow rate is determined by external factors controlling the rate of mass flow ($\mathrm{d}m/\mathrm{d}t$).

The most effective means of heat transfer by mass transport is through the latent heat of vaporization of the fluid. *Latent heat of vaporization is the amount of energy required to vaporize 1 kg of the fluid that is already at its boiling point.* It is denoted by Λ. To vaporize 1 kg of water, 2.4 MJ of heat is required whereas to heat water through 100°C, only 0.42 MJ is required. A schematic of a water-cooled system is shown in Figure 13.21. The heat source is a power semiconductor device that dissipates heat. This heats up the water in the chamber connected to the heat source to a temperature T_1. When the water temperature reaches boiling point, then T_1

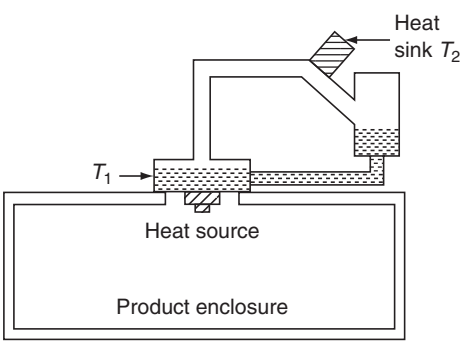


Figure 13.21 | Water-cooled system.

is at 100°C which is maintained as constant. Water takes the heat from the source; evaporates and carries the heat energy corresponding to the latent heat of vaporization with it; and transfers to the sink which is at temperature T_2. The sink is maintained at temperature T_2 by means of cold water or ice or peltier junctions. After giving the heat to the sink, the water vapor again converts back to liquid form and flows into the collection chamber which then links to the evaporation chamber as shown. Thus, it can be observed that heat taken from the heat source at T_1 is carried to wherever the vapor condenses at T_2. The associated heat flow rate is given by

$$P_{\mathrm{m}} = \frac{\mathrm{d}m}{\mathrm{d}t}\Lambda \tag{13.43}$$

where $\mathrm{d}m/\mathrm{d}t$ is the rate at which fluid is being evaporated and Λ is the latent heat of vaporization of the fluid. The associated thermal resistance is given as

$$R_{\mathrm{m}} = \frac{T_1 - T_2}{P_{\mathrm{m}}} = \frac{T_1 - T_2}{(\mathrm{d}m/\mathrm{d}t)\Lambda} \tag{13.44}$$

13.6 General Notes on Heat Transfer

Mixed Heat Transfer Mechanisms

It should be noted that even though the various heat transfer mechanisms have been discussed in an isolated manner wherein one mode of heat transfer is decoupled from the other mechanism, in reality some or all of the heat transfer mechanisms may be active simultaneously in a thermal system. For example, in the case of the mass transport mechanism discussed in the previous section, the pipe through which the fluid flows is at a higher temperature with respect to the external ambient. This implies that there will exist convective and radiative heat energy-transfer mechanisms from the hot pipe to the ambient. Therefore, one will have to consider all the effects together. The general rate of heat energy flow is given by Eq. (13.1). Let P_{c}, P_{v}, P_{r} and P_{m} be the heat power flow through the conductive, convective, radiative and mass transport mechanisms, respectively. If $R_{\theta\mathrm{c}}$, $R_{\theta\mathrm{v}}$, $R_{\theta\mathrm{r}}$ and $R_{\theta\mathrm{m}}$ are the thermal resistance for the conductive, convective, radiative and mass transport mechanisms, respectively, then

$$P_{\mathrm{tot}} = P_{\mathrm{c}} + P_{\mathrm{v}} + P_{\mathrm{r}} + P_{\mathrm{m}} \tag{13.45}$$

$$P_{\mathrm{tot}} = \frac{\Delta T_{\mathrm{c}}}{R_{\theta\mathrm{c}}} + \frac{\Delta T_{\mathrm{v}}}{R_{\theta\mathrm{v}}} + \frac{\Delta T_{\mathrm{r}}}{R_{\theta\mathrm{r}}} + \frac{\Delta T_{\mathrm{m}}}{R_{\theta\mathrm{m}}} \tag{13.46}$$

Equation (13.45) follows from the conservation of energy principle. In Eq. (13.46), ΔT_{c}, ΔT_{v}, ΔT_{r} and ΔT_{m} are the temperature differences between the hot body and the lower temperature references for the conductive, convective, radiative and mass transport mechanisms. If the temperature differences are same for all the mechanisms, then Eq. (13.46) can be written as

$$P_{\mathrm{tot}} = \left(\frac{1}{R_{\theta\mathrm{c}}} + \frac{1}{R_{\theta\mathrm{v}}} + \frac{1}{R_{\theta\mathrm{r}}} + \frac{1}{R_{\theta\mathrm{m}}} \right) \Delta T \tag{13.47}$$

or if the area perpendicular to the heat power flow is same for all mechanisms, then the power flow in terms of the thermal coefficients for the various heat transfer mechanisms can be written as

$$P_{\mathrm{tot}} = (h_{\theta\mathrm{c}} + h_{\theta\mathrm{v}} + h_{\theta\mathrm{r}} + h_{\theta\mathrm{m}}) A \Delta T \tag{13.48}$$

$$P_{\text{tot}} = \left(\frac{k}{\Delta x} + \frac{k\mathcal{N}}{X} + 4\sigma\varepsilon_{\text{eff}}(1-\phi)\left(\frac{T_1+T_2}{2}\right)^3 + \frac{(\mathrm{d}m/\mathrm{d}t)s}{A} \right) A\Delta T \tag{13.49}$$

Thus, the total power flow including all the heat transfer effects is given as

$$P_{\text{tot}} = h_{\text{eq}} A\Delta T \tag{13.50}$$

where

$$h_{\text{eq}} = \left(\frac{k}{\Delta x} + \frac{k\mathcal{N}}{X} + 4\sigma\varepsilon_{\text{eff}}(1-\phi)\left(\frac{T_1+T_2}{2}\right)^3 + \frac{(\mathrm{d}m/\mathrm{d}t)s}{A} \right)$$

It should be noted that Eq. (13.50) is an algebraic summation and sometimes some of the power flow mechanism may have a canceling effect on the power flow by other mechanism. However, Eq. (13.49) would in general give a conservative value of the equivalent thermal coefficient that can be used for design purposes.

Thermal Resistance of Varying Cross-Sections

Consider the heat transfer by conduction mechanism. The cross-section area A that is orthogonal to the heat power flow direction is considered constant till now. However, there are cases where A varies with distance. The power flow equation is given as

$$P = kA\frac{\Delta T}{\Delta x} = kA\frac{\mathrm{d}T}{\mathrm{d}x}$$

$$\mathrm{d}T = \frac{P}{kA}\mathrm{d}x \tag{13.51}$$

Integrating Eq. (13.51) between temperature limits T_2 and T_1 and the length from 0 to L, one obtains

$$\int_{T_2}^{T_1} \mathrm{d}T = \int_0^L \frac{P}{kA}\mathrm{d}x$$

$$\Delta T = P\int_0^L \frac{\mathrm{d}x}{kA} \tag{13.52}$$

$$R_\theta = \int_0^L \frac{\mathrm{d}x}{kA}$$

Equation (13.52) is a general equation where A can vary along the path having changing cross-sections. Further, if the solid material is a eutectic combination of different materials, then k also varies along the path.

Consider a system as shown in Figure 13.22. The heat source is mounted at the center of a solid block that is either rectangular or circular as shown. The heat power flow radiates in all directions from the center as shown. At a greater distance from the heat source, the cross-section area is larger as the radius is larger. The cross-section area at a radius r is given as $2\pi rb$. Thus,

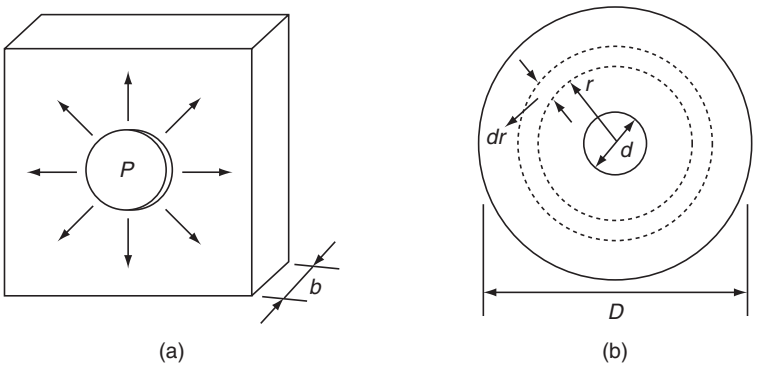


Figure 13.22 Non-uniform cross-section area: (a) Rectangular block; (b) increasing radius in the direction of power flow.

$$R_\theta = \int_{d/2}^{D/2} \frac{\mathrm{d}r}{k(2\pi rb)}$$
$$= \frac{1}{2\pi kb} \ln r\Big|_{d/2}^{D/2} = \frac{1}{2\pi kb} \ln\left(\frac{D}{d}\right) \tag{13.53}$$

It should also be noted that in the case of a flat plate, if both the surfaces are exposed to the ambient, then the heat power flows to the ambient on both sides of the flat plate. In such cases, area A is taken as twice the area of one side of the flat plate to account for the power flow from both surfaces for any heat transfer mechanism.

Mounting of Power Devices on Heat Sink

Special care should be taken during mounting of the power semiconductor devices on the heat sink to minimize the thermal resistance from the case of the device to the heat sink, that is, $R_{\theta cs}$. The surface of the power device case and the surface of the heat sink have micro-undulations which reduce the contact between the two surfaces as shown in Figure 13.23(a). As a consequence, the contact thermal resistance will increase thereby increasing the case-to-sink thermal resistance. To improve the contact between the device case surface and the heat-sink surface, a thin layer of heat-sink paste or heat-sink compound is applied over the heat-sink surface to fill up the micro-undulations as shown in Figure 13.23(b). This improves the contact between the two surfaces. Care must be taken to ensure that the heat-sink paste should be applied in a very thin layer just enough to fill up the micro-undulations. Thicker layer of heat-sink paste will also increase the overall thermal resistance.

In some applications, more than one device needs to be mounted on the same heat sink. The case of the power devices are generally connected to one of the device leads, that is, the collector is also the case in the case of the BJT, the drain is also the case in the case of MOSFETs, etc. In such cases, to prevent shorting of the terminals of the devices through the heat sink, each power device is mounted on the heat sink with a thin sheet of insulating mica layer interposed between the device and the heat-sink surface as shown in Figure 13.23(c).

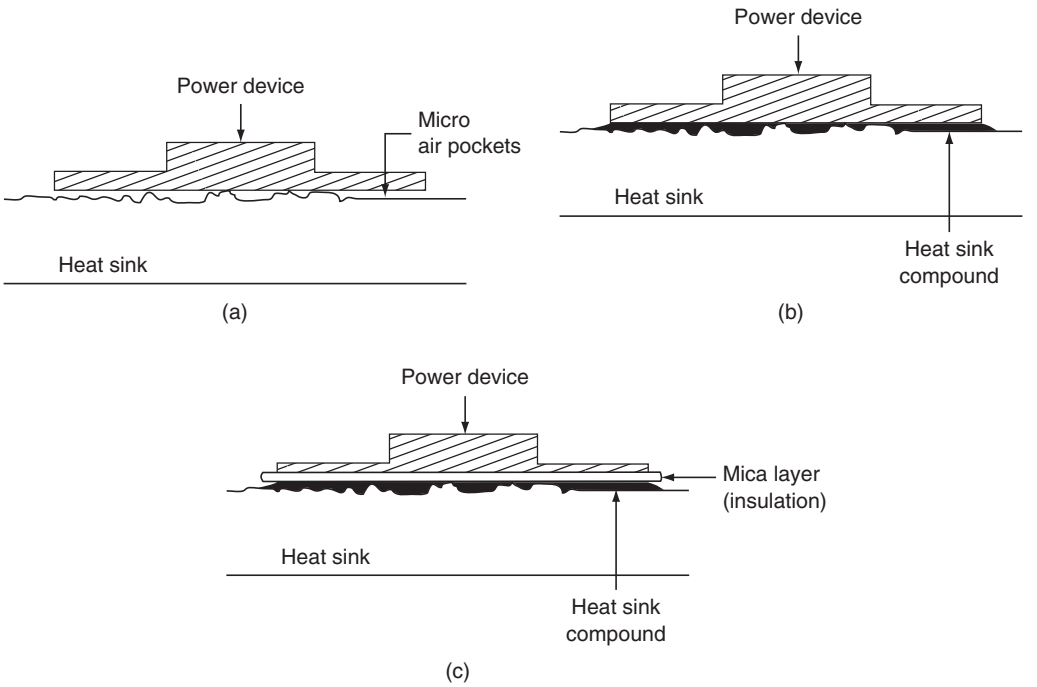


Figure 13.23 | Power device mounted on heat sink: (a) Without heat-sink compound; (b) after applying heat-sink compound; (c) after applying heat-sink compound and inserting a layer of mica for insulation.

Table 13.3 | Case-to-sink thermal resistance for TO-220 and TO-3 device package

	$R_{\theta cs}$ in °K/W			
	Direct mount		Mount with mica insulator	
Device package	Without heat-sink compound	With heat-sink compound	Without heat-sink compound	With heat-sink compound
TO-220	1.4	0.5	3	1.4
TO-3	0.6	0.1	1.25	0.5

Table 13.3 gives typical value of the case-to-sink thermal resistance for two of the popular device packages with a torque of 0.4–0.6 Nm applied on the device mounting screws.

Transient Model for Heat Transfer

In most of the applications, the steady-state model is sufficient to design or select the heat sink to achieve proper heat transfer from the device to the external ambient. However, in some applications, the load is pulsed and the dissipated power is also of a pulse nature. In such cases, the heat-sink size and volume may

be further optimized if the dynamics of the temperature rise and fall are taken into account. For the thermal domain, the mass of the solid or the fluid acts as the capacitance which stores the thermal energy by virtue of the raised temperature. The heat energy from the heat source is transferred into two sinks: (a) the external ambient which is the infinite sink and (b) the mass of the medium through which the heat transfers. The energy transferred into the medium is called the stored thermal energy. Thus,

$$E_{\text{input}} = E_{\text{ext-ambient}} + E_{\text{stored}}$$

and the power flow is given as

$$P = \frac{\mathrm{d}E_{\text{input}}}{\mathrm{d}t} = \frac{\mathrm{d}E_{\text{ext-ambient}}}{\mathrm{d}t} + \frac{\mathrm{d}E_{\text{stored}}}{\mathrm{d}t} \tag{13.54}$$

$$P = \underbrace{kA\frac{\mathrm{d}T}{\mathrm{d}x}}_{\text{Steady-state part}} + \underbrace{mc_{\text{p}}\frac{\mathrm{d}T}{\mathrm{d}t}}_{\text{Transient part}} \tag{13.55}$$

where m is the mass of the medium in which heat is stored; c_{p} the specific heat of the medium in which heat is stored. The term mc_{p} is equivalent to the capacitance in the electrical circuit. Equation (13.55) gives the dynamic equation that defines the heat energy flow rate.

Best Method of Heat Transfer

There is no single best method of heat transfer for removing heat from the power device to the external ambient. Each of the heat transfer mechanism has its inherent advantages and disadvantages. However, in general, the lower the thermal resistance required, more expensive will be the heat transfer system. The following illustration in Figure 13.24 gives a general guideline on the heat transfer method that can be used for various heat source power levels.

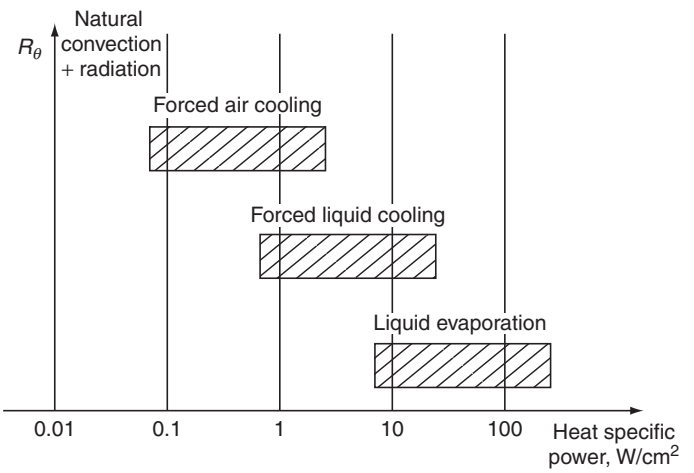


Figure 13.24 | Relative effectiveness of different heat transfer mechanisms.

Referring to Figure 13.24, it can be observed that for low specific heat power flows, the thermal resistance obtainable from natural or free convection and radiation is sufficient to maintain the junction temperature of the semiconductor devices within the safe temperature limits. For very high specific heat power, the mass transport with liquid evaporation (similar to Figure 13.21) is the most effective in reducing the thermal resistance.

CONCLUDING REMARKS

In the chapters discussed till now, the focus has been essentially on addressing electrical parameters and stresses. However, this is only half the job. Unless the thermal stress aspect is addressed, the design of power electronics systems is never complete. This chapter addresses this important issue of keeping the thermal stresses in the power semiconductor devices within safe limits.

The power semiconductor devices generate heat due to the conduction loss and the switching loss as discussed in Chapter 1. Therefore the power semiconductor devices behave as heat sources. If the heat generated by the devices are not conducted away to the ambient, the temperature of the device junction will increase and damage the device. In order to ensure that the junction temperature of the semiconductor devices are maintained within permissible limits, the heat sinks are used to conduct away the heat from the junction to the ambient. Therefore, the heat-sink design would mean the estimation of the required sink to ambient thermal resistance that is needed to transfer the heat from the device case to the ambient such that the junction temperature is within safe limits.

Heat sinks are available commercially in the form of aluminum extrusions. There are numerous aluminum extrusion profiles that are available that enables a designer to meet both the thermal aspects and the product form requirements. The next section gives few tutorial tasks that should necessarily be performed by prospective designers in order to achieve a certain measure of skill and confidence in handling the thermal stress aspects.

TUTORIAL EXERCISES

1. **Heat source circuit setup:** In order to study the thermal aspects, one needs a controllable heat source to experiment with the heat transfer mechanisms. The circuit shown in Figure 13.25a consists of a BJT that is operated in the linear region. The base source V_b and the collector supply V_c are dc sources. The base drive resistor R_b is a variable resistor. Adjust either V_b or R_b to allow a pre-determined base current, I_b. This will result in a collector current depending on the h_{FE} of the BJT. The base current is set such that the operating point of the BJT is as shown in Figure 13.25b. The power dissipation in the BJT is given by the product of I_C and V_{CE}. By adjusting I_b, any desired power dissipation can be achieved. Therefore, the BJT operated in the linear region becomes a controllable heat source.

   Mode of implementation: The above circuit can be studied by

   a. Hardware bread-boarding

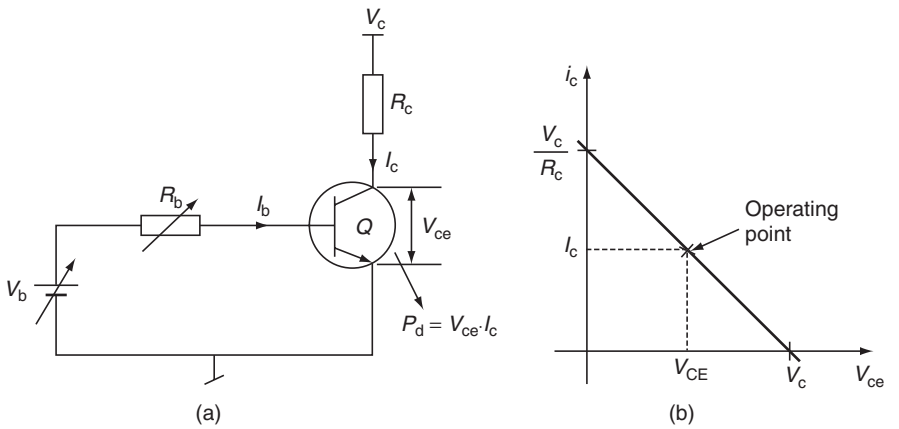


Figure 13.25 | BJT test circuit.

Tasks for study:

(a) Set V_b to be a particular dc value. Measure the collector current and the collector–emitter voltage across transistor for different values of R_b. Tabulate the BJT collector current, base current, emitter current and collector–emitter voltage values. Calculate the product of I_C and V_{CE} which is the power dissipation of the BJT.

(b) For every setting of I_b, allow the case temperature of the BJT to reach the equilibrium state. (Allow the circuit to operate for around 15 min to reach steady state.) Also measure the room temperature or the ambient temperature at every power-dissipation setting. Tabulate the above readings.

(The above two tasks are for understanding the heat power source.)

2. **Thermal resistance measurements:** Mount the heat source BJT, as described in Tutorial 1, at the center of a flat rectangular aluminum plate of known dimensions.

Mode of implementation:

a. Hardware bread-boarding

Tasks for study:

(a) Set the heat source power to a particular value. Allow the system to reach a steady state in terms of the BJT case temperature for horizontal position of the aluminum plate. From the datasheet of the BJT, obtain the junction-to-case thermal resistance value.

(b) Measure and tabulate the heat source power, the BJT case temperature, the temperature of the aluminum plate near the heat source and also at the edges. Measure and tabulate the room temperature also.

(c) From the above data, calculate the case-to-sink thermal resistance and sink-to-ambient thermal resistance.

(d) Estimate the junction temperature of the BJT heat source.

(e) Repeat tasks (b) to (d) by mounting the BJT heat source with and without heat-sink compound. Tabulate and compare the thermal resistances.

(f) Theoretically calculate the aluminum sink to ambient thermal resistance due to convection and radiation heat transfer mechanisms. Compare the theoretical estimation with the experimental results.

(g) Now repeat tasks (b) to (f) for vertical orientation of the rectangular plate. The rectangular plate can be placed vertical in two ways, (i) smaller side vertical and (ii) longer side vertical. Conduct the experiment for both orientations.

(h) Repeat tasks (b) to (g) after including a fan for providing forced convection. Perform the experiment for different fan speeds (air velocities), tabulate and plot the sink-to-ambient thermal resistance versus fan speeds or air velocities.

FILL IN THE BLANKS

1. The junction temperature should not exceed the _____ junction temperature which is _____ for most devices.

2. Conduction is a process of heat transfer wherein the heat flows through _____ from a higher temperature zone to a lower temperature zone.

3. Convection is the process of heat transfer from solids by means of _____.

4. Radiation is a process of heat transfer, wherein the heat flows out from the hot body without the need for a heat transport _____.

5. Convection is an _____ process where as mass transport is a controlled process where the discharge rate of the fluid can be controlled resulting in _____ heat transfer from one region to another.

6. With reference to the heat flow, the _____ difference is considered equivalent to the potential difference of electrical circuit.

7. The rate of heat flow or power flow is considered equivalent to the _____ of electrical circuit.

8. There is no concept of _____ in the thermal domain.

9. The thermal resistance is not a _____ element, it is only a notional element that indicates the conductivity of heat through the medium.

10. _____ is sometimes referred to as the ohm's law for thermal domain.

11. If there is a heat flow through a solid, then it implies a temperature _____ along the length of the solid slab in the direction of the _____.

12. Thermal conductivity is a _____ property and is the characteristic of the solid slab material.

13. The thermal resistance of silver is _____ than that of steel.

14. The convection heat transfer mode is composed of following two mechanisms: (a) _____ and (b) _____ of the fluid.

15. The flow induced by diffusion which arises from density differences caused by temperature variations in the fluid is called _____ convection.

16. The flow caused by external means such as fan, pump or winds is called _____ convection.

17. The Nusselt number is a _____ quantity.

18. Nusselt number is a function of the _____ of fluid flow, _____ of the hot solid and the properties of the fluid medium.

19. Nusselt number also depends on the type of convection, whether _____ or _____.

20. Nusselt number in the case if free convection is estimated using the _____ number.

21. Nusselt number in the case of forced convection is estimated using the _____ number.

22. In forced convection, a _____ fluid flow is applied along the surface of the solid hot body to improve the convective heat transfer.

23. The convective thermal resistance is dependent on the _____ of the hot solid surface in the case of free convection.

24. The convective thermal resistance is _____ of the orientation of the hot solid surface in the case of the forced convection.

25. A substance at a finite temperature will emit energy called _____ radiation that is transported by _____ waves.

26. The transfer of heat energy by radiation occurs most efficiently in _____.

27. The shielding factor gives a measure of the _____ of the radiative heat transfer from a hot body due to the presence of another body in the vicinity.

28. For a single plate or two parallel plates, the shielding factor is _____.

29. The emittance is the energy emitted relative to that of a _____.

30. The value of emittance is less than _____.

31. Heat transfer by mass transport involves the _____ interaction.

32. The _____ is the heat source and the heat from the source is transferred to the _____ which takes away the heat energy.

33. The flow rate of the fluid can be used for _____ the heat energy transfer.

34. The heat energy from the heat source is transferred into two sinks, (a) the external ambient which is the _____ sink and (b) the mass of the medium through which the heat transfers.

35. The energy transferred into the mass of the medium is called the _____ thermal energy.

36. The term _____ is equivalent to the capacitance in the electrical circuit.

37. Mass transport by evaporation gives the lowest _____ to the heat flow among the four heat transfer mechanisms.

DESCRIPTIVE QUESTIONS

1. What are the four types of heat transfer mechanisms?

2. Write short notes on: (a) conduction, (b) convection, (c) radiation and (d) mass transport.

3. Write short notes on: (a) specific heat flow rate, (b) thermal resistance, (c) thermal resistivity, (d) thermal coefficient and (e) thermal conductivity.

4. What is the heat flow rate equation? Discuss.

5. Write short notes on: (a) Nusselt number, (b) Rayleigh number and (c) Reynolds number.

6. Write short notes on: (a) free convection and (b) forced convection.

7. Explain why there is a measure of uncertainty in the estimation of the convective thermal resistance.

8. How is Rayleigh number evaluated for a fluid?

9. How is Reynolds number evaluated for a fluid?

10. A flat plate that is oriented vertically has a *lower* thermal resistance as compared to the same flat plate that is oriented horizontally.

11. Discuss the determination of Nusselt number for few standard geometries of the hot solid for free convection.

12. Discuss the determination of Nusselt number for few standard geometries of the hot solid for forced convection.

13. Write short notes on: (a) reflectance, (b) absorptance and (c) transmittance.
14. How does heat transfer by mass transport differ from heat transfer by convection?
15. Discuss the mounting considerations for the power devices on the heat sink.
16. Discuss the choice of the heat transfer mechanism for any given application.

PROBLEMS

1. Calculate thermal resistance to the flow of power through a solid square slab of 100 cm length wherein
   a. the slab is made up of aluminum with thickness of 10 mm.
   b. the slab is made up of glass with thickness of 5 mm.
   c. the slab is made up of steel with thickness of 2 mm.
2. A MOSFET that is used in a dc–dc converter is dissipating 50 W. The thermal resistance to conduction from the junction to the case is 0.5°K/W and the thermal resistance to conduction from the case to the heat sink is 1.5°K/W. If the ambient temperature in the neighborhood of the heat sink is 50°C, then calculate the thermal resistance requirement for the heat sink if the junction temperature does not exceed (a) 100°C, (b) 120°C.
3. A MOSFET that is used in an inverter is dissipating 20 W. The thermal resistance to conduction from the junction to the case is 0.5°K/W and the thermal resistance to conduction from the case to the heat sink is 1.5°K/W. The heat sink used ensures that the device junction temperature does not exceed 100°C. If the ambient temperature in the neighborhood of the heat sink is 50°C, then estimate the device case temperature.
4. Consider a hot flat square aluminum plate measuring 100 cm on each side that is placed vertically. The temperature at a distance from the flat plate is 45°C. The temperature of the square aluminum plate is 70°C. Heat flows to the ambient from both faces of the aluminum plate. Calculate the heat flow rate in watts from the plate to the external ambient by convection through the air.
5. For Problem 4, if a fan is placed at the bottom of the vertical plate such that air flows up along both sides of the vertical plate at a velocity of 2 m/s, then estimate the convective thermal resistance.
6. Consider the problem given in Problem 13.3 which is illustrated in Figure 13.16. The specifications are modified as follows. The solid cylindrical heat sink has a diameter of 22 cm and a height of 11 cm. The ambient temperature is 20°C and the space within the enclosure is at 45°C. The cylinder is at a temperature of 100°C due to the dissipation from the heat source, that is, the MOSFET mounted underneath the cylindrical heat sink. The dimensions of the cuboidal steel enclosure are 100 cm × 100 cm × 10 cm. There is conductive heat flow from the MOSFET source to the heat sink; and from the heat sink to the ambient there is both convective and radiative heat flow. There is conductive, convective and radiative heat flow from the enclosure to the ambient. Considering all these heat flows, calculate the power flow from the heat source (MOSFET) that is required to maintain the heat-sink temperature at 100°C under the following conditions:

a. Almost still air environment for the system.
b. Forced convection with a fan mounted that provides an air flow of 3 m/s for the heat sink.

7. Consider two parallel plates each having an area of 1 m^2. Let the first plate be made up of anodized aluminum and kept at a temperature of 373°K. Let the second plate be made up of rough surfaced aluminum and the temperature of the plate be 300°K. Find the radiation heat flow from first plate to the second plate.

8. Derive the thermal resistance due to mass transport mechanism of heat transfer by evaporation of water.

9. Consider a circular heat source of diameter 10 mm mounted at the center of a circular aluminum plate of diameter 100 cm and thickness 15 mm. What is the temperature gradient between the circumference of the circular plate and the heat source?

ANSWERS

Fill in the Blanks

1. maximum-rated; 150°C
2. solids
3. fluids
4. medium
5. uncontrolled; controlled
6. temperature
7. current flow
8. dissipation
9. dissipative
10. $\Delta T = P \cdot R_\theta$
11. gradient; heat flow
12. transport
13. lower
14. diffusion; bulk motion
15. free
16. forced
17. dimensionless
18. velocity; geometry
19. free; forced
20. Rayleigh
21. Reynolds
22. controlled
23. orientation
24. independent
25. thermal; electromagnetic
26. vacuum
27. shading
28. zero
29. black body
30. unity
31. solid–fluid
32. solid; fluid
33. controlling
34. infinite
35. stored
36. $(m \cdot c_p)$
37. thermal resistance

CHAPTER 14

Reliability Modeling

Learning Objectives

After reading this chapter, you will be able to:

- understand the need for reliability.
- understand the concepts of reliability and reliability modeling.
- apportion the reliability and specify reliability for various modules.
- design for reliability.

There are many constraints that need to be considered in the overall design of a system. One such constraint is a set of performance requirements that needs to be addressed. Apart from performance, cost is another important constraint that needs to be addressed in the design of the system. The third aspect is the life constraint that considers aspects related to the life of the product. In most cases, the three constraints may be contradictory to each other. For example, a high-accuracy performance and a low-cost product may be contradictory to each other. Likewise, long life and low cost are also contradictory by nature. In general, any product may be said to be a compromise among the three primary constraints: (a) performance, (b) cost and (c) life (Figure 14.1).

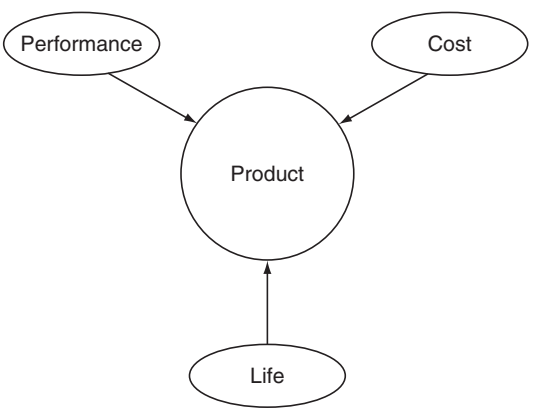


Figure 14.1 The three primary constraints in designing a product.

A concept of the product that achieves an *optimal* compromise among the three primary constraints of performance, cost and life is said to be a *good* design. The previous chapters discuss about aspects related to the performance of the power electronic systems. However, a power electronic system design or analysis is not only about performance. This chapter focuses on the life aspect and discusses issues related to reliability of the products and the process of integrating reliability requirements into the design of the system.

One can appreciate the significance of designing for reliability when one visualizes the power electronic system or any product development process. A typical product development process is illustrated in Figure 14.2. The first task is to define the system in terms of the specifications. The specifications should incorporate the requirements of all the three primary constraints, that is, performance, life and cost. Specification for reliability is in terms of failure rate or mean time to failure (MTTF). The specification process will result in a set of target specifications that needs to be achieved by the design and development process. The system is designed to meet the performance specifications followed by validation of the design through simulation and bread-boarding. This is followed by the reliability modeling of the system and reliability prediction to estimate the life of the product. This is an important process as the guarantees, warranties and service infrastructural requirements are decided by achieving this constraint. This is followed by cost analysis and cost optimization. There is a feedback loop on completing the process related to each of the primary constraints. The system design has to be modified or fine-tuned to realize all the specification requirements.

The previous chapters focused on the performance design process whereas this chapter will focus on the reliability modeling issues. The next chapter will discuss reliability prediction in detail and the method of integrating the reliability requirements into the design process to achieve better design cycle time.

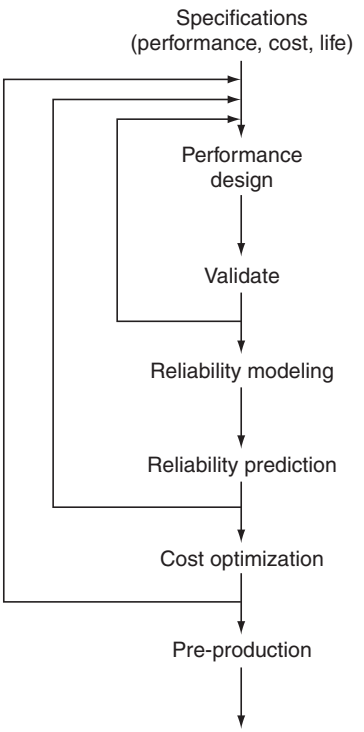


Figure 14.2 Design process for a product.

14.1 Reliability Concepts

The problem of assuring reliability has many facets including equipment design, control of quality during production, acceptance inspection, field trials, life testing and design modifications. The various electronic components are subjected to two types of stresses: (a) electrical stress and (b) temperature stress. The electrical stress may be in terms of voltage-withstanding capability of the device or the current-handling capability of the device. The temperature stress are due to the internal dissipation which will raise the junction temperature of the semiconductor devices. The component fails due to the application of these two types of stresses during operation. However, the failure of a component is a random process as the exact cause and reason for the failure is uncertain. Therefore, *reliability is defined as the probability that an item will perform its intended function for a specified period of time under the stated conditions*. Thus, it can be seen from this definition that reliability encompasses both performance (intended function) and life (specified period of time) requirements.

Since reliability has been defined as a probability, the quantitative analysis is based on probability-based tools. For the purpose of reliability discussion in this chapter, let T be a continuous random variable that denotes the time at which failure of an item occurs. Therefore, according to the definition of the continuous random variable, there must exist a function $f(t)$, called the failure-time probability density function. Therefore, one has

$$f(t) \geq 0, \quad 0 < t < \infty$$

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = \int_{0}^{\infty} f(t)\mathrm{d}t = 1$$

$$\mathrm{P}(a < T < b) = \int_{a}^{b} f(t)\mathrm{d}t$$

$$\mathrm{P}(T < t) = \int_{0}^{t} f(t)\mathrm{d}t = F(t)$$

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = f(t)$$

One should note that $F(t) = \mathrm{P}(T < t)$ denotes the probability that the random variable T takes on values less than t, that is, it denotes that failure occurs in the time interval $(0, t)$. Evidently, $1 - F(t)$ denotes that the failure does not occur in the time interval $(0, t)$. This is called reliability $R(t)$. Thus the reliability of an item can be mathematically defined as

$$R(t) = 1 - F(t) \tag{14.1}$$

Failure Rate

Reliability has been described as quality in the time dimension. It is classically defined as the probability that an item will perform satisfactorily for a specified period of time under a stated set of use conditions. From a functional point of view, for an item to be reliable, it must do more than meet an initial factory performance or quality specification and must also operate satisfactorily for an acceptable period of time in field application for which it is intended. Determining reliability, therefore, involves understanding of several concepts which relate to the four elements in the classical definition, that is, probability, performance, time and stated

use conditions. Among such concepts is that of the failure rate which can vary as a function of time as discussed earlier. A failure rate is a measurement of the number of malfunctions occurring per unit time.

Referring to Figure 14.3, let E_1 be an event that indicates the occurrence of failure in the time interval $(0, t)$. Then $\overline{E_1}$ is the event that failure does not occur in the time interval $(0, t)$. Let E_2 denote the event that the failure occurs in the interval $(t, t + \Delta t)$.

Clearly, E_1 and E_2 are mutually exclusive events which are indicated in Figure 14.3(b). It is also evident from the figure that $\overline{E_1}$ and E_2 are not mutually exclusive events. One is interested in the probability of occurrence of a failure in the interval $(t, t + \Delta t)$. Thus, one has to evaluate the probability of occurrence of event E_2 on the condition that E_1 has not occurred or in other words, the probability of occurrence of event E_2 on the condition that $\overline{E_1}$ has occurred. Thus,

$$\mathrm{P}(E_2 / \overline{E_1}) = \frac{\mathrm{P}(E_2 \cap \overline{E_1})}{\mathrm{P}(\overline{E_1})} \tag{14.2}$$

From Figure 14.3(b), it is evident that event E_2 is wholly contained in $\overline{E_1}$. Thus,

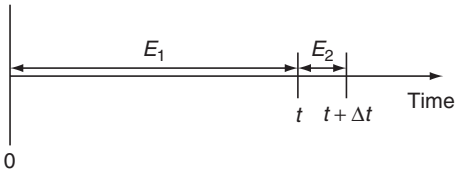
$$\mathrm{P}(E_2 \cap \overline{E_1}) = \mathrm{P}(E_2) \tag{14.3}$$

Substituting Eq. (14.3) in Eq. (14.2), one obtains

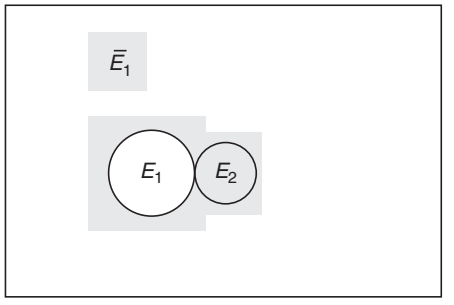
$$\mathrm{P}(E_2 / \overline{E_1}) = \frac{\mathrm{P}(E_2)}{\mathrm{P}(\overline{E_1})} \tag{14.4}$$

$$\mathrm{P}(E_2) = \mathrm{P}(t < T < t + \Delta t) = F(t + \Delta t) - F(t) \tag{14.5}$$

$$\mathrm{P}(\overline{E_1}) = 1 - \mathrm{P}(E_1) = 1 - \mathrm{P}(T < t) = 1 - F(t) = R(t) \tag{14.6}$$



(a)



(b)

Figure 14.3 | Events to describe the failure-time distribution.

Substituting Eqs. (14.5) and (14.6) in Eq. (14.4), the probability that an item will fail in the interval $(t, t+\Delta t)$ on the condition that the component has survived till time t is given by

$$\mathrm{P}(E_2 / \overline{E_1}) = \frac{F(t+\Delta t) - F(t)}{R(t)} \tag{14.7}$$

The probability in the case of continuous sample spaces is the area under the probability density function within the given interval. Referring to Eq. (14.7), dividing it by the base Δt, the average rate of failure in the interval t to $t+\Delta t$ on the condition that the item has survived till time t is given by

$$\frac{\mathrm{P}(E_2 / \overline{E_1})}{\Delta t} = \frac{F(t+\Delta t) - F(t)}{\Delta(t)}\left(\frac{1}{R(t)}\right) \tag{14.8}$$

Taking the limit as $\Delta t \to 0$, the instantaneous failure rate or simply the failure rate denoted by $Z(t)$ is given by

$$Z(t) = \lim_{\Delta t \to 0}\left[\frac{F(t+\Delta t) - F(t)}{R(t)}\left(\frac{1}{R(t)}\right)\right] \tag{14.9}$$

$$Z(t) = \frac{F'(t)}{R(t)} \tag{14.10}$$

where

$$F'(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t} = f(t)$$

$$Z(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)} \tag{14.11}$$

One should note that in several literature, the *failure rate* $Z(t)$ is also referred to as the *hazard rate*. Here, the terms failure rate and the hazard rate will be used interchangeably.

Failure-Time Probability Density Function

From the definition of reliability as indicated by Eq. (14.1), the following relationship can be obtained by differentiating Eq. (14.1):

$$R'(t) = -F'(t) \tag{14.12}$$

Using Eq. (14.10) in Eq. (14.12), one obtains

$$Z(t) = -\frac{R'(t)}{R(t)} = -\frac{\mathrm{d}[\ln R(t)]}{\mathrm{d}t} \tag{14.13}$$

Thus integrating Eq. (14.13), one obtains

$$\int_0^t Z(t)\mathrm{d}t = -\ln R(t) \tag{14.14}$$

Equation (14.14) can be re-written as

$$R(t) = e^{-\int_0^t Z(t)\mathrm{d}t} \tag{14.15}$$

From Eq. (14.10) it is evident that $f(t) = Z(t)R(t)$. Thus, from Eq. (14.15), the failure-time probability density function is given by

$$f(t) = Z(t)\, e^{-\int_0^t Z(t)\mathrm{d}t} \tag{14.16}$$

Equation (14.16) is an interesting relationship which shows that the failure-time probability density function is completely determined by the failure rate or the hazard rate $Z(t)$ for a particular process. Consider, for example, that the failure rate for some system is a constant λ. The failure-time probability density function can be uniquely determined by using Eq. (14.16). Thus,

$$f(t) = \lambda e^{-\int_0^t \lambda \mathrm{d}t} = \lambda e^{-\lambda t} \tag{14.17}$$

Clearly, it is observed that when the failure rate is constant, the resulting failure-time probability density function has the characteristics of an exponential distribution. In some cases, one may encounter a distribution where the failure rates may not be a constant but can increase with time or decrease with time. Therefore, one can use the following general model for the failure rate,

$$Z(t) = \lambda\beta t^{\beta-1} \quad (t > 0) \tag{14.18}$$

It is evident from Eq. (14.18) that if $\beta = 1$, then the failure rate is a constant. If $\beta < 1$, then the failure rate decreases with time. On the other hand, if $\beta > 1$, the failure rate increases with time. Using the failure rate model of Eq. (14.18) in Eq. (14.16), the failure-time probability density function is given as

$$f(t) = \lambda\beta t^{\beta-1} e^{-\lambda t^{\beta}} \tag{14.19}$$

The distribution given in Eq. (14.19) is called the Weibull distribution where the exponential distribution is a special case at $\beta = 1$.

To show the variation in the failure rate, separate consideration is given to three discrete periods when viewing the failure characteristics of a product or item over its life span. These periods, which are indicated in Figure 14.4, are

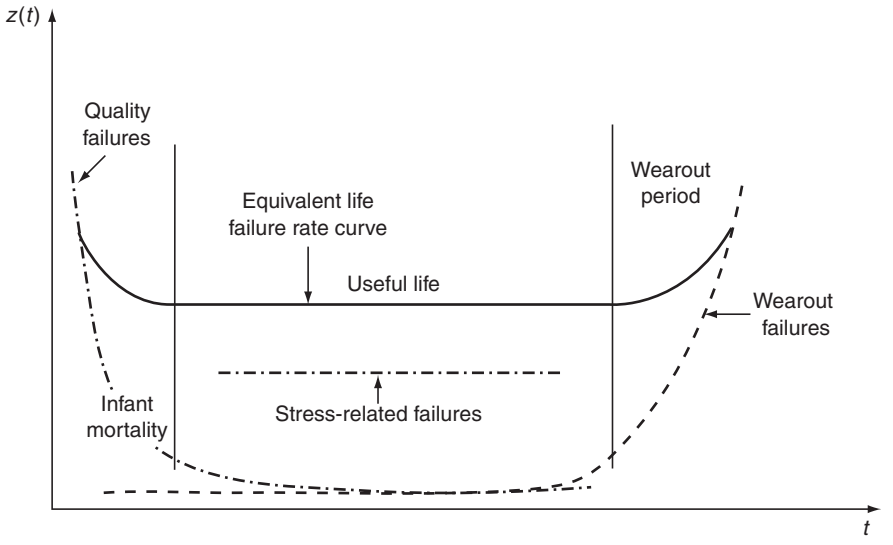


Figure 14.4 | Failure rate curve showing the failure rate during the life of an electronic equipment.

1. infant mortality period;
2. useful life period;
3. wear out period.

Infant Mortality Period

Initially, the item population exhibits a high failure rate. This failure rate decreases rapidly during this first period, often called the infant mortality, burn-in or debugging period. The failure rate then stabilizes when the weak units have died out. The failures in this period are generally related to quality. It may be caused by a number of things: gross built-in flaws due to faulty workmanship (manufacturing deviations from the design intent), transportation damage or installation errors. This initial failure rate is unusually pronounced in new equipment. Many manufacturers provide a burn-in period for their product prior to delivery as this helps to eliminate a high portion of the initial failures and assists in establishing a high level of operational reliability. Examples of early failures are:

1. poor welds or seals;
2. poor solder joints;
3. poor connections;
4. dirt or contamination on surfaces or in materials;
5. chemical impurities in metal or insulation;
6. voids, cracks, thin spots in insulation or protective coatings;
7. incorrect positioning of parts.

Many of these early failures can be prevented by improving the control over the manufacturing process. Sometimes, improvements in design or materials are required to increase the tolerance for these manufacturing deviations, but fundamentally these failures reflect the manufacturability of the component or product and the control of the manufacturing process. Consequently, these early failures would show up during

1. in-process and final tests;
2. process audits;
3. life tests;
4. environmental test.

Useful Life Period

The item population, after having been burned-in, reaches its lowest failure rate level which is normally characterized by a relatively constant failure rate. This period is characterized by the occurrence of *stress-related failures*. As indicated in Figure 14.4, the stress-related failures have a constant failure rate. Thus the exponential failure-time probability density function is widely used as a mathematical model to approximate this time period. This period is the interval which is given the most weightage in reliability design and is the most significant period for reliability prediction and assessment activities.

Wear Out Period

The third and the final life period occurs when the item population reaches the point where the failure rate starts to increase noticeably. This point is identified as the end of the useful life and the start of wear out or ageing. Beyond this point on the time axis, the failure rate increases rapidly. When the hardware failure rate due to wear out becomes unacceptably high, replacement or repair of the item should be made. Replacement schedules are based on the recognition of this failure rate.

Wear out failures as shown in Figure 14.4 are primarily due to deterioration of the strength of the device as a consequence of operation and exposure to environmental fluctuations. Deterioration results from a number of familiar chemical and physical phenomena:

1. corrosion or oxidation;
2. insulation breakdown or leakage;
3. frictional wear or fatigue;
4. shrinkage and cracking in plastics.

Mean Time Between Failure (MTBF), Mean Time to Failure (MTTF)

The period of interest during the life cycle of a product or a system is the useful life period. During this period, the hazard rate or the failure rate is constant. Therefore, the failure-time probability density function of the product or the system can be modeled as an exponential distribution. If the failure rate is λ, then

$$f(t) = \lambda e^{-\lambda t} \tag{14.20}$$

where λ is the failure rate which is expressed as the number of failures that occur per unit time. It implies that the expected value of the random variable, that is, the expected time to failure of the item is $1/\lambda$. Clearly, the MTTF of an item is the reciprocal of the failure rate, that is

$$\text{MTTF} = \frac{1}{\lambda} = \frac{1}{\text{Failure rate}} \tag{14.21}$$

Figure 14.5 gives a visualization of the MTTF. Referring to Figure 14.5, consider that a new item starts functioning from time $t = 0$. The failure-time probability density function starts decreasing as shown. At time $t = 1/\lambda$, the item is expected to fail. If at that instant the item is replaced by a new item, then its failure-time probability density function again starts to decrease from λ and so on.

MTBF is used for repairable systems and for the case of non-repairable items the MTTF is used. Thus,

$$\text{MTBF} = \text{MTTF} = \frac{1}{\lambda}$$

and λ is the failure rate in number of failures per unit time.

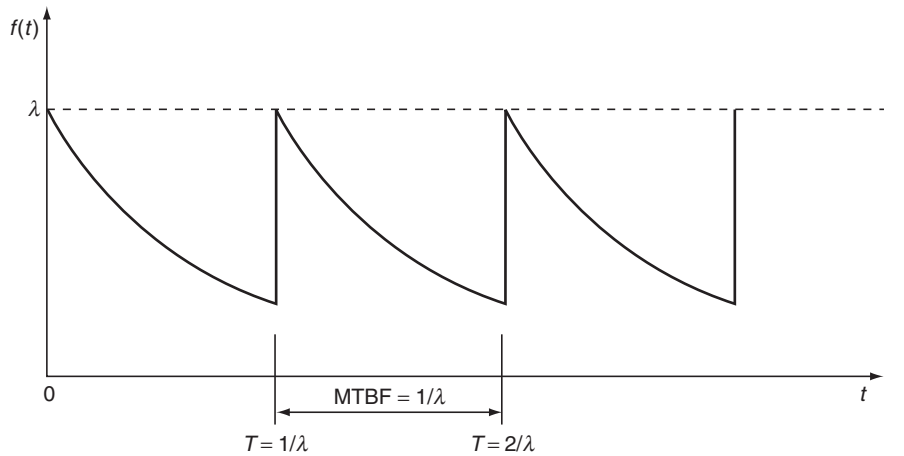


Figure 14.5 | A visualization of the MTBF.

It should be understood that as reliability is a probabilistic value. If the MTTF (MTBF) for a product is 10 years, say, then it means that the reliability is $R(t) = e^{-\lambda t} = e^{-t/\text{MTTF}} = e^{-0.1t}$. This means that at the end of 10 years, the reliability would have become $e^{-1} = 0.366$. This implies that at the end of 10 years, 63.6% of the products released in the market would have come back for servicing. It may not be a good idea to give a 10-year warranty for such a product. On the other hand, after 1 year, the reliability is $e^{-0.1} = 0.904$ implying that only 10% of the products originally released in the market would be back for servicing. Statistically, it may be better to give a one-year warranty for the product that is designed to have a MTTF of 10 years.

14.2 Modeling for Reliability

Prior to the use of reliability analysis techniques, it is essential to model any given system in a manner that is conducive for quantitative analysis. Therefore, a method is needed to reflect the reliability connectivity of the many part types having different failure rates that would normally make up a complex system. This is accomplished by establishing a relationship between equipment reliability and individual item failure rates.

In practice, a system is frequently represented as a network in which the system components are connected together either in series, parallel, meshed or a combination of these. It is vital that the relationship between the system and its network model be thoroughly understood before considering the analytical techniques that can be used to evaluate the reliability of these networks. It must be appreciated that the actual system and the reliability network used to model the system may not necessarily have the same topological structure. This consideration involves the essential point that the analyst must be fully familiar with the requirements of the system and be able to phrase these requirements in a form which can be quantitatively assessed.

Two fundamental reliability network models that are frequently encountered are the series systems and the parallel systems. In a series system, from the reliability point of view, the system will fail even if one of the items of the series system fails. On the other hand, in a parallel system, the system will fail only if all the items of the parallel system fail.

Series System

Consider a system consisting of n independent components as shown in Figure 14.6. The n components are connected in series from the reliability point of view. This arrangement implies that all components must work to ensure system success.

Let $R_1(t)$, $R_2(t)$, …, $R_n(t)$ be the respective probabilities that the components do not fail upto time t. Let $R(t)$ be the reliability of the entire system. Let T_1 be the random variable that indicates the time to failure of component 1; T_2 be the random variable that indicates the time to failure of component 2 and so on. Let T be the random variable that indicates the time to failure of the overall series-connected system. Then

$$\text{P}(T > t) = \text{P}(T_1 > t) \,\&\, \text{P}(T_2 > t) \,\&\, \cdots \,\&\, \text{P}(T_n > t) \tag{14.22}$$

As the operations of the n components are independent, Eq. (14.22) can be re-written as

$$R(t) = R_1(t)R_2(t)\ldots R_n(t) \tag{14.23}$$

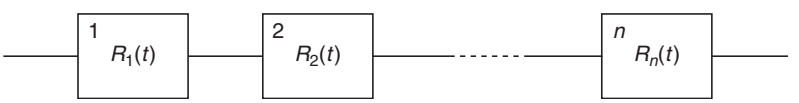


Figure 14.6 | Series configuration.

Thus the reliability of the series-connected system, where $R_1(t)$, $R_2(t)$, … $R_n(t)$ are the respective individual component reliabilities, is given by

$$R(t) = \prod_{i=1}^{n} R_i(t) \tag{14.24}$$

In the case of components with constant failure rate, the reliability of the series system is given by

$$R(t) = e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} \cdots\cdots e^{-\lambda_n t} = e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)t} \tag{14.25}$$

where λ_1, λ_2, …, λ_n are the respective individual component failure rates. It is evident from Eq. (14.25) that if the individual components have failure-time probability densities that are exponential, the resulting series system also has an exponential failure-time probability distribution with an equivalent failure rate of $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

As $\lambda_i > 0$ for all i, it is evident that the equivalent failure rate of the series-connected system is higher than the component having the largest failure rate.

MTBF of the equivalent series system is given by

$$\text{MTBF} = \frac{1}{\lambda} = \frac{1}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \tag{14.26}$$

Problem 14.1

A system design requires 200 identical components in series. If the overall reliability must not be less than 0.99, what is the minimum reliability of each component?

Solution

Here, $0.99 = R^{200}$

Thus, $R = 0.99^{1/200} = 0.99995$

Parallel System

Consider a system consisting of n independent components connected in parallel, from the reliability point of view, as shown in Figure 14.7. In this case, the system fails only if all the components fail. Let $R_i(t)$ be the reliability of the ith component and $F_i(t)$ be the unreliability or the probability that the component fails in time t of the ith component. Let $R(t)$ and $F(t)$ be the reliability and unreliability of the entire parallel system respectively. Then, the system fails only if (component 1 fails) & (component 2 fails) & … so on. Thus,

$$F(t) = F_1(t)F_2(t) \ \ldots \ F_n(t) \tag{14.27}$$

Equation (14.27) can be re-written as

$$1 - R(t) = [1 - R_1(t)][1 - R_2(t)]\cdots\cdots[1 - R_n(t)] \tag{14.28}$$

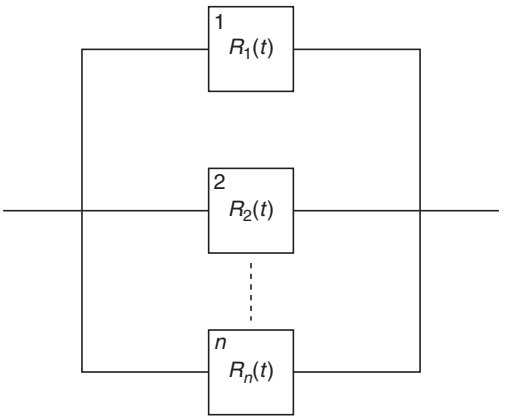


Figure 14.7 | Parallel configuration.

The reliability of the parallel system can be deduced from Eq. (14.28) and is given by

$$R(t) = 1 - \prod_{i=1}^{n} [1 - R_i(t)] = 1 - \prod_{i=1}^{n} F_i(t) \tag{14.29}$$

It is evident from Eq. (14.29) that even if the individual components have failure-time distributions that are exponential in nature, the failure-time distribution of the equivalent parallel system is not exponential in nature. Therefore, if one has to evaluate the equivalent failure rate or the MTBF for the parallel system, one has to use the following relations:

$$Z(t) = \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)} \tag{14.30}$$

$$\text{MTBF} = E(T) = \int_0^\infty t \cdot f(t) \mathrm{d}t \tag{14.31}$$

One should note that the parallel system is nothing but a redundant system. As all the components are working, it is called full-on redundant system. Further, as the whole system works successfully (survives) even with just one component working, the parallel system described till now is called a *full-on*, *single-survivor parallel redundant system*. In later sections, different types of redundant configurations will be discussed.

In redundant systems, it is common to encounter identical individual components in the parallel system. Thus the reliabilities of each of the system would be identical and equal to $e^{-\lambda t}$ where λ is the failure rate of each component of the parallel system. In such a case, for n components in parallel, using Eq. (14.29) one obtains

$$R(t) = 1 - (1 - e^{-\lambda t})^n \tag{14.32}$$

Using the binomial expansion, Eq. (14.32) can be expanded in series form as

$$R(t) = nC_1 e^{-\lambda t} - nC_2 e^{-2\lambda t} + \cdots + (-1)^{n-1} e^{-n\lambda t} \tag{14.33}$$

Using the relationship that the failure-time probability density function $f(t)$ is related to the reliability as $f(t) = -R'(t)$, the failure-time probability density function $f(t)$ can be obtained from Eq. (14.33). Thus,

$$f(t) = nC_1\lambda e^{-\lambda t} - nC_2(2\lambda)e^{-2\lambda t} + \cdots + (-1)^{n-1}(n\lambda)e^{-n\lambda t} \tag{14.34}$$

Thus, Eq. (14.34) describes the failure-time probability density function for the parallel redundant system with identical individual components. To obtain the MTBF for a parallel system with identical components, one can start from the definition of the mean of the random variable which is

$$\text{MTBF} = E(T) = \mu = \int_0^\infty t \cdot f(t)\mathrm{d}t \tag{14.35}$$

Substituting Eq. (14.34) in Eq. (14.35), one obtains

$$\mu = nC_1\int_0^\infty t \cdot \lambda e^{-\lambda t}\mathrm{d}t - nC_2\int_0^\infty t \cdot 2\lambda e^{-2\lambda t}\mathrm{d}t + \cdots + (-1)^{n-1}\int_0^\infty t \cdot n\lambda e^{-n\lambda t}\mathrm{d}t \tag{14.36}$$

Equation (14.36) reduces to

$$\mu = nC_1\left(\frac{1}{\lambda}\right) - nC_2\left(\frac{1}{2\lambda}\right) + \cdots + (-1)^{n-1}\left(\frac{1}{n\lambda}\right) \tag{14.37}$$

Expanding the combination terms and substituting $n = 1, 2, 3, \ldots$ in Eq. (14.37), the MTBF for a parallel system with identical individual components is given by

$$\text{MTBF} = \mu = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \tag{14.38}$$

Problem 14.2

Consider the reliability block diagram shown in Figure 14.8. Evaluate the reliability of the system.

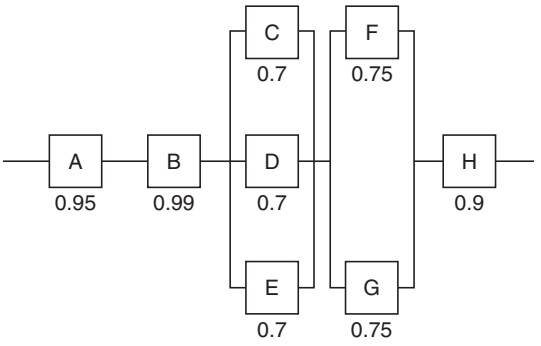


Figure 14.8 | Problem 14.2.

Solution

The reliability of the system is given by

$$R(t) = (0.95)(0.99)[1 - (1 - 0.7)^3][1 - (1 - 0.75)^2](0.9) = 0.772$$

Problem 14.3

Consider the full-bridge rectifier–capacitor filter circuit as shown in Figure 14.9. The input is a sinusoidal waveform given by 100sin(ωt). The capacitor has been so designed such that the output voltage ripple is 10 V for full-wave rectification operation and 20 V for half-wave rectification operation. Let q_d be the probability of failure of the diodes and q_c be the probability of failure of the capacitor. Draw the reliability block diagram and find the reliability of the system if the minimum output voltage required is (a) 85 V, (b) 75 V.

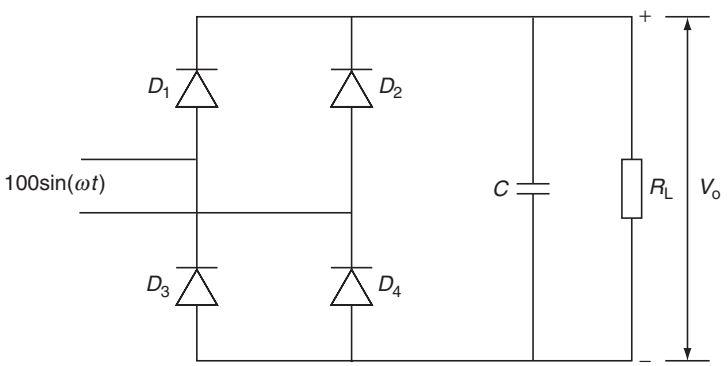


Figure 14.9 | Problem 14.3.

Solution

Let E_{Di} denote the event indicating the occurrence of failure of the ith diode and E_C denote the event indicating the occurrence of failure of the capacitor. Let $R(t)$ and $F(t)$ be the reliability and the unreliability of the system. Then,

Case (a)

Here it is required that the output voltage should be at least 85 V for proper operation. If the intended function of at least 85 V output is not maintained, then it is considered as system failure. Thus,

System fails IF (E_{D1} and/or E_{D2} and/or E_{D3} and/or E_{D4}) occurs and/or (E_C) occurs

The reliability block diagram is shown in Figure 14.10(a). From the reliability block diagram, the reliability is given by

$$R(t) = (1 - q_d)^4 (1 - q_c)$$

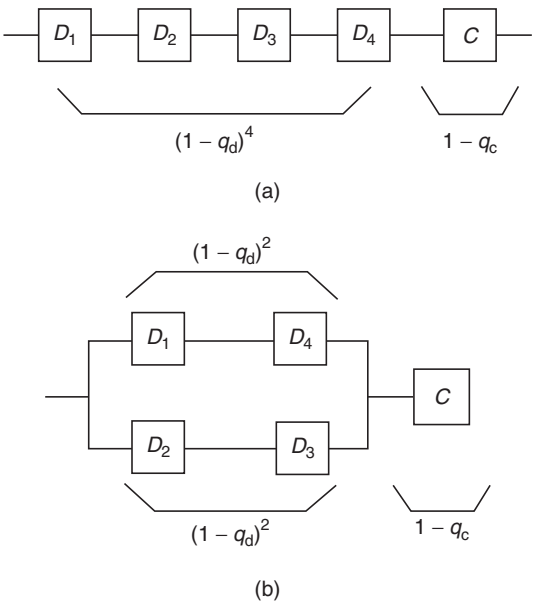


Figure 14.10 | (a) For Case I; (b) for Case II.

Case (b)

Here it is required that the output voltage should be at least 75 V for proper operation. In this case, for example if D_1 open circuits, then half-wave rectification operation occurs through D_2 and D_3. As the output voltage ripple is 20 V under this mode of operation, it is within the limits of proper operation. Therefore, any failures or combination of diode failures that leads to half-wave rectification operation is also proper operation. Thus,

System works IF $\overline{E_{D1}}$ & $\overline{E_{D4}}$ occurs and/or $\overline{E_{D2}}$ & $\overline{E_{D3}}$ occurs and $\overline{E_C}$ occurs

The reliability block diagram is as shown in Figure 14.10(b). From the reliability block diagram, the reliability is given by

$$R(t) = \{1-[1-(1-q_d)^2]^2\}(1-q_c)$$

Problem 14.4

Consider the half-wave rectifiers shown in Figure 14.11. All diodes are identical with an open-circuit failure probability of q_o and a short-circuit failure probability of q_s. If the intended function of the circuit is in obtaining a half-wave rectified waveform at the output, then evaluate the reliabilities of the three circuits through reliability block diagrams. If $q_o = 0.05$ and $q_s = 0.3$, then evaluate the reliabilities for the three circuits.

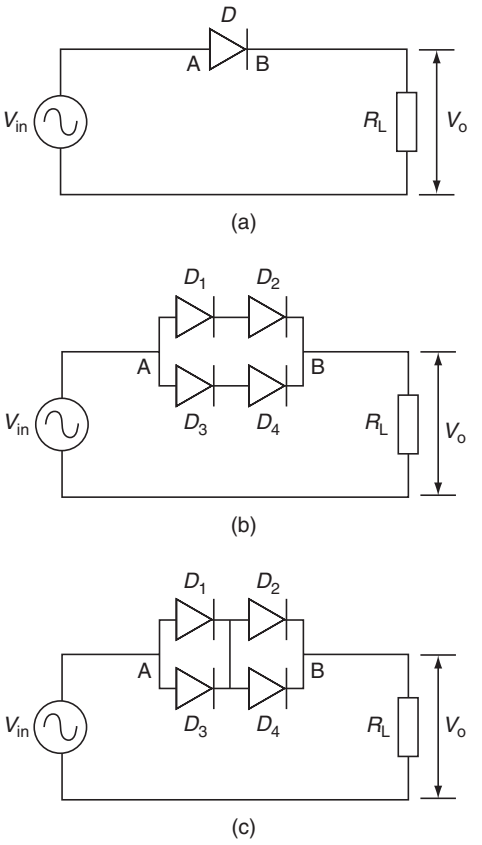


Figure 14.11 | Problem 14.4.

Solution

If the system has to fail then there should either exist a short or an open between the points A and B in the case of all three circuits. In all three cases let E_{Di} represent an event where the diode D_i fails.

Case I

Here the system fails if the diode fails either in the short-circuit mode or in the open-circuit mode. Thus, the reliability is given by

$$R(t) = 1 - (q_s + q_o)$$

as the failure of the diode in any one of the modes (short or open) are mutually exclusive. Substituting the values of q_s and q_o, one obtains

$$R(t) = 0.65$$

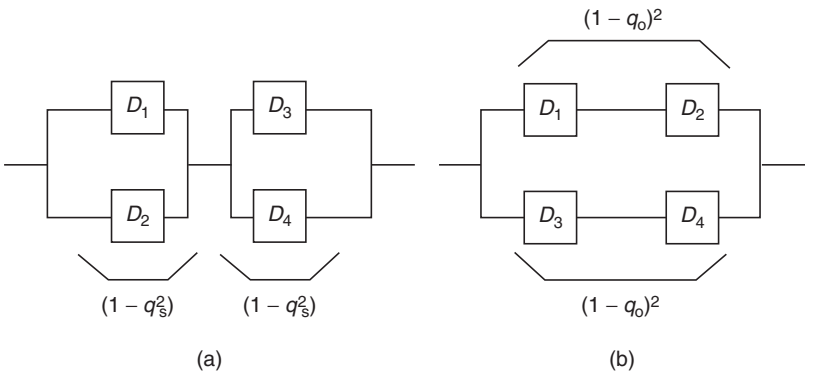


Figure 14.12 | Case II: (a) short-circuit failure mode; (b) open-circuit failure mode.

Case II

Here one has to consider the two failure modes between the points A and B, that is, the short-circuit failure mode and the open-circuit failure mode as shown in Figure 14.12.

Short-circuit failure mode: The system fails by a short between points A and B

$$\text{IF } (E_{D1} \,\&\, E_{D2}) \text{ occur and/or } (E_{D3} \,\&\, E_{D4}) \text{ occur}$$

This results in the short-circuit failure mode reliability block diagram as shown in Figure 14.12(a). From the reliability block diagram the unreliability is given by

$$F_s(t) = 1 - (1 - q_s^2)^2$$

Open-circuit failure mode: The system fails by a open between points A and B

$$\text{IF } (E_{D1} \text{ and/or } E_{D2}) \text{ occur } \& \ (E_{D3} \text{ and/or } E_{D4}) \text{ occur}$$

The open-circuit reliability block diagram for this failure mode is indicated in Figure 14.12(b). The unreliability is given by

$$F_o(t) = [1 - (1 - q_o)^2]^2$$

The short-circuit failure mode and the open-circuit failure mode are mutually exclusive; thus the unreliability and reliability of the total system are given as

$$F(t) = F_s(t) + F_o(t) = 1 - (1 - q_s^2)^2 + [1 - (1 - q_o)^2]^2$$
$$R(t) = (1 - q_s^2)^2 - [1 - (1 - q_o)^2]^2$$

Substituting the values of q_o and q_s in the above reliability equation, one obtains

$$R(t) = 0.8186$$

Case III

Here again, one has to consider the two failure modes between the points A and B, that is, the short-circuit failure mode and the open-circuit failure mode.

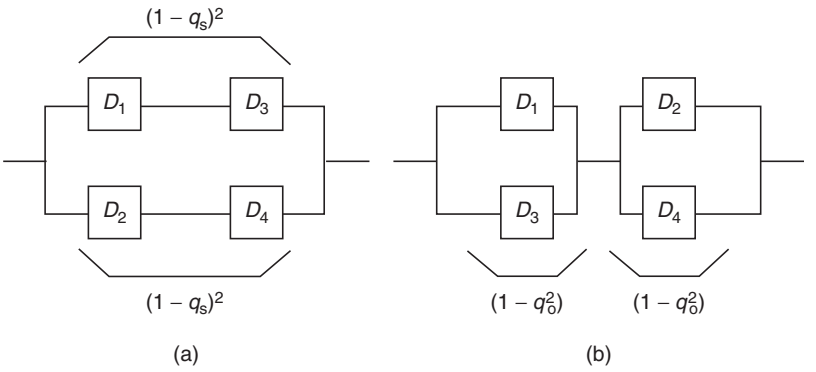


Figure 14.13 | Case III: (a) short-circuit failure mode; (b) open-circuit failure mode.

Short-circuit failure mode: The system fails by a short between points A and B

$$\text{IF } (E_{D1} \text{ and/or } E_{D3}) \text{ occur \& } (E_{D2} \text{ and/or } E_{D4}) \text{ occur}$$

From the short-circuit failure mode reliability block diagram [Figure 14.13(a)] the unreliability is given by

$$F_s(t) = [1-(1-q_s)^2]^2$$

Open-circuit failure mode: The system fails by a open between points A and B

$$\text{IF } (E_{D1} \text{ \& } E_{D3}) \text{ occur and/or } (E_{D2} \text{ \& } E_{D4}) \text{ occur}$$

From the open-circuit failure mode reliability block diagram [Figure 14.13(b)] the unreliability is given by

$$F_o(t) = 1-(1-q_o^2)^2$$

The short-circuit failure mode and the open-circuit failure mode are mutually exclusive, thus the unreliability and reliability of the total system are given as

$$F(t) = F_o(t) + F_s(t) = 1-(1-q_o^2)^2 + [1-(1-q_s)^2]^2$$
$$R(t) = (1-q_o^2)^2 - [1-(1-q_s)^2]^2$$

Substituting the values of q_o and q_s in the above reliability equation, one obtains

$$R(t) = 0.735$$

Referring to Problem 14.4, it is seen that both circuits of Figure14.11(b) and (c) are more reliable than the circuit of Figure 14.11(a). Between Case II and Case III, it is seen that Case II is more immune to short-circuit failures, that is, this configuration is preferred where the short-circuit failure probability of the diodes is higher than the open-circuit failure probability. Case III is more immune to open-circuit failures, that is, this configuration is preferred where the open-circuit failure probability of the diodes are higher than the short-circuit failure probability.

Mesh System

Many systems do not have the simple structure as expressed by systems exhibiting series and parallel structures. They may have complex operational logic. Additional modeling and evaluation techniques are necessary to determine the reliability of such systems. A typical system not having a series/parallel structure is the bridge-type network as shown in Figure 14.14. A visual inspection of the network shown in Figure 14.14 indicates that the components are not connected in a simple series/parallel arrangement. There are a number of techniques available for solving this type of network such as the conditional probability approach, cut and tie set analysis, tree diagrams, logic diagrams and connection matrix techniques. It should be noted that most of the methods are very similar in concept. The essential difference between them lies in the formal presentation or logic of the method and not the underlying concept.

Consider the system shown in Figure 14.14 in which success requires that at least one path exists from point X to point Y. Here component E is a bi-directional component. If each component has a failure probability of q, then it is required to evolve a method to evaluate the reliability of the system.

Let E_A, E_B, E_C, E_D, E_E be the events that represent the failures of components A, B, C, D and E, respectively. One can assume that the failures of each component are independent of the failures of the other components.

On visual inspection, it is evident that there are four possible paths from X to Y. They are AC, BD, AED and BEC. Thus the reliability block diagram for the system is as shown in Figure 14.15.

This is now a parallel system where the system works if any one of the paths exist. The probability that the system works or the reliability is given by

$$R(t) = \mathrm{P}(\overline{E_A}\ \&\ \overline{E_C})\ \text{and/or}\ \mathrm{P}(\overline{E_B}\ \&\ \overline{E_D})\ \text{and/or}\ \mathrm{P}(\overline{E_A}\ \&\ \overline{E_D}\ \&\ \overline{E_E})\ \text{and/or}\ \mathrm{P}(\overline{E_B}\ \&\ \overline{E_C}\ \&\ \overline{E_E})$$

Using the third axiom of probability and substituting "q" as the failure probability for each component, one obtains

$$R(t) = 2p^2 + 2p^3 - 5p^4 + 2p^5$$

where $p = 1 - q$. However, the above method is not strictly a parallel system as discussed in section "Parallel System". In a true parallel configuration, the failure of each of the path is independent of the failures of the other paths. In the case of the parallel configuration of Figure 14.15, the parallel paths are not independent as the failure of one component influences the failure or success of other paths. Therefore, the results obtained by this method can at times be dubious.

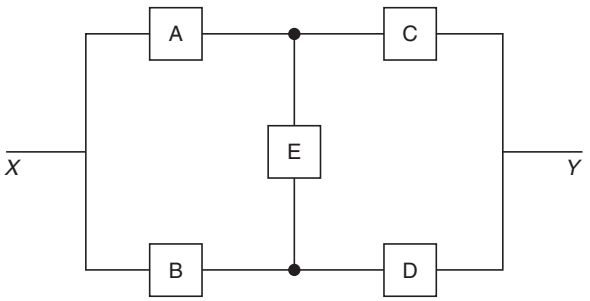


Figure 14.14 | Mesh configuration.

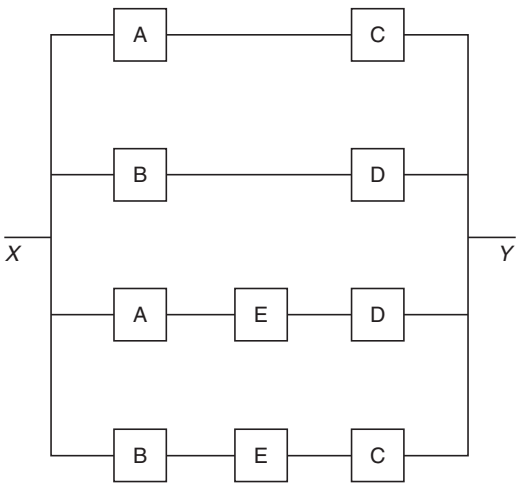


Figure 14.15 | Multiple path equivalent of the mesh configuration.

Conditional Probability Approach

In this approach, complex systems are reduced sequentially into sub-system structures that are connected in series/parallel and then to recombine these sub-systems using the conditional probability method. The basic concept of this approach is as follows:

P(system success or failure) = P(system success or failure if component C is good) × P(C is good) + P(system success or failure if component C is bad) × P(C is bad)

To apply the conditional probability to the problem of Figure 14.14, it is first necessary to choose the reference component which will be considered "good" and considered "bad". Any of the components A to E may be chosen but some choices will result in smaller sub-branches. Here component E is chosen as the reference component. When E is good, the system reduces to a simple parallel–series system as shown in Figure 14.15(b). When E is bad, the system reduces to a simple series–parallel system as shown in Figure 14.16(c). These two sub-systems are mutually exclusive because E_E and $\overline{E_E}$ are mutually exclusive and further they are collectively exhaustive. Therefore, the system reliability is given by

$$R_s(t) = \text{P(system works if E works)} \times \text{P(E works)} + \text{P(system works if E fails)} \times \text{P(E fails)}$$

$$R_s(t) = \text{P}(\overline{E_S} / \overline{E_E}) \cdot \text{P}(\overline{E_E}) + \text{P}(\overline{E_S} / E_E) \cdot \text{P}(E_E) \tag{14.39}$$

where E_S is the event that the system fails. Referring to Figure 14.16,

$$\text{P}(\overline{E_S} / \overline{E_E}) = (1 - F_A F_B)(1 - F_C F_D) \tag{14.40}$$

$$\text{P}(\overline{E_S} / E_E) = 1 - [(1 - R_A R_C)(1 - R_B R_D)] \tag{14.41}$$

Substituting Eqs. (14.40) and (14.41) in Eq. (14.39), one obtains

$$R_s(t) = (1 - F_A F_B)(1 - F_C F_D) R_E + \{1 - [(1 - R_A R_C)(1 - R_B R_D)]\} F_E \tag{14.42}$$

Equation (14.42) can be simplified to obtain

$$R(t) = 2p^2 + 2p^3 - 5p^4 + 2p^5$$

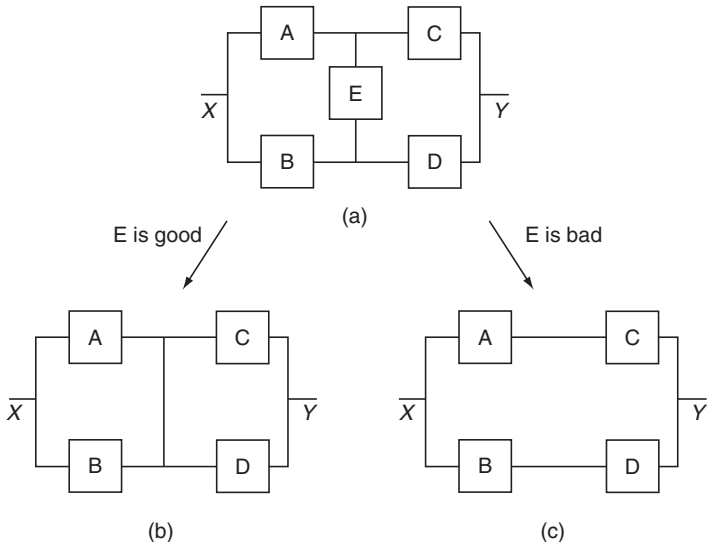


Figure 14.16 | Conditional probability approach.

where $p = 1 - q$. One should note that in the above example, only one level of sub-division was performed. But in very complex systems, one or more sub-systems may need further sub-division before a series/parallel structure is obtained. This is only an extension of the technique that was discussed above since each time a sub-division is made, the two sub-divisions must be recombined using the conditional probability approach starting at the lowest hierarchical level.

14.3 Redundant Systems

To cope with technological developments, systems have been compelled to expand both in size and complexity at a rather rapid rate. Of equivalent importance to the need for this growth has been the simultaneous need for greater system reliability. As industrial systems become more complex, the philosophy of simply increasing the reliability by making parts more reliable became unrealistic due to technological limits. The only answer to such situations has been the design of redundant systems, sometimes termed as fault tolerant systems.

Redundancy exists when one or more components of a system fail and the system can continue to perform its intended function satisfactorily. The system may adapt to component failures by performing the required function with the remaining components or it may switch-in an available spare component to take the place of the failed one. The components used to take the place of the failed ones might be functionally redundant, that is, not identical in form to the failed unit but capable of compensating for the failed function; for example, one processor taking over the role of a different type of processor, a VHF transceiver taking over the function of a UHF transceiver, etc., or it could be an exact duplicate of the failed component.

There are many different redundancy strategies. Some work in conjunction with maintenance and/or corrective action performed periodically. Some are applied when no maintenance or corrective action

can be performed. Each has its advantages (like reliability gain) and disadvantages (like system weight, cost, volume, number of duplicative elements, etc.). In the following discussion on redundant systems, systems without maintenance are considered.

Full-On Redundancy: Multiple-Survivor Sub-System

This type of redundancy is a more general case of the simple parallel system discussed earlier. The system consists of n fully energized parallel components but requires a minimum of m working components for the system to work. The m-survivor system is depicted in Figure 14.17. For this m-survivor system,

1. $R_c(t)$ is the probability of a component of the system functioning for a period of time t.
2. n is the number of components in the parallel system.
3. Functioning of each of the n components is independent.
4. There are only two possible states of the components either working (success) or failed (failure) state.

The minimum number of components required to work for the system to function satisfactorily is given by m. It is evident from points (1) to (4) that the choice of m working components out of n components is a Bernoulli process. Thus, binomial distribution is applicable to the choice of m working components. One should note that at the start, that is, at $t = 0$, all n units are assumed to be functional.

The reliability $R(t)$ of the system is the probability that the random variable T takes on values greater than t. This happens only if at least m components of the n components are working. Thus,

$$R(t) = \mathrm{P}(T > t) = \text{Probability that at least } m \text{ units must function}$$

Let X be a random variable denoting the number of functional units. Now it is of interest to find the probability that the random variable X takes on values greater than or equal to m for satisfactory system functioning.

$$R(t) = \mathrm{P}(T > t) = \mathrm{P}(X \geq m) \tag{14.43}$$

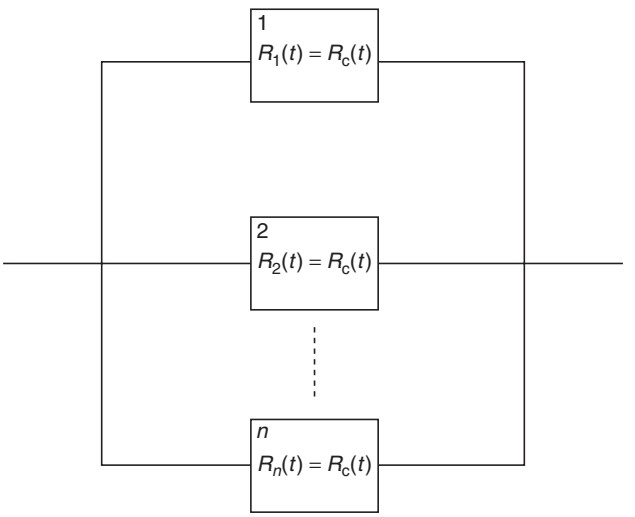


Figure 14.17 The m-survivor full-on redundant system.

From the Binomial distribution,

$$\mathrm{P}(X \geq m) = \sum_{k=m}^{n} nC_k [R_\mathrm{c}]^k [1 - R_\mathrm{c}]^{n-k} \tag{14.44}$$

Alternatively consider Y to be a random variable denoting the number of non-functional units. Then for the system to be functioning, there should be less than or equal to $n - m$ failed units. Thus,

$$\mathrm{P}(X \geq m) = \mathrm{P}(Y \leq n - m) \tag{14.45}$$

Again from the Binomial distribution,

$$\mathrm{P}(X \geq m) = \mathrm{P}(Y \leq n - m) = \sum_{k=0}^{n-m} nC_k [1 - R_\mathrm{c}]^k [R_\mathrm{c}]^{n-k} \tag{14.46}$$

where $R_\mathrm{c} = e^{-\lambda \mathrm{t}}$ is the reliability of each of the components of the redundant system.

MTTF or MTBF for Full-On m-Survivor Redundancy

From the basic definition of the mean or expectation of the random variable, the MTBF is given by,

$$\mathrm{MTBF} = \int_0^\infty t \cdot f(t) \mathrm{d}t \tag{14.47}$$

It is known that

$$f(t) = F'(t) = -R'(t) \tag{14.48}$$

Substituting Eq. (14.48) in Eq. (14.47), one obtains

$$\mathrm{MTBF} = -\int_0^\infty t \cdot R'(t) \mathrm{d}t \tag{14.49}$$

Integrating by parts and making use of the fact that $R(\infty) = 0$ (because the probability that the system will work for infinite time is zero), one obtains

$$\mathrm{MTBF} = \int_0^\infty R(t) \mathrm{d}t \tag{14.50}$$

One should note that Eq. (14.50) is applicable for any distribution. For the m-survivor case where all the components have exponential distribution with identical failure rates, the MTBF is given by

$$\mathrm{MTBF} = \frac{1}{\lambda} \left(\frac{1}{m} + \frac{1}{m+1} + \cdots + \frac{1}{n} \right) \tag{14.51}$$

One should note at this point that for $m = 1$, the system becomes a full-on, single-survivor redundant system which is the simple parallel network discussed earlier.

Problem 14.5

An array of signal processors in a parallel computing system performs a specific function. A total of 100 processors make up the array. The failure rate of each processor is 0.0005 failures/h. The system can perform satisfactorily as long as at least 90 processors are operating.

(a) What is the probability that the system will operate satisfactorily over a 100-h period, given that all components were operable at the start?
(b) What is the MTTF of the system?
(c) What would the above reliability characteristics be if a minimum of 95 processors must operate?

Solution

$$R_c = e^{-\lambda t} = e^{-0.0005\times 100} = 0.95$$

Here $n = 100$, $m = 90$, $1 - R_c = 0.05$. For $n > 20$ and $p < 0.05$, one can also use the Poisson approximation to the Binomial distribution. Thus,

$$R(t) = \sum_{k=0}^{10} 100C_k (0.05)^k (0.95)^{100-k} = \sum_{k=0}^{10} \frac{\alpha^k e^{-\alpha}}{k!}$$

where $\alpha = np = (100)(0.05) = 5$.

(a) $$R(t) = \sum_{k=0}^{10} \frac{5^k e^{-5}}{k!} = 0.0067 + 0.0337 + 0.0842 + 0.1403 + 0.1754 + 0.17546 + 0.1462 + 0.1044 + 0.06527 + 0.03626 + 0.01813 = 0.986$$

(b) $$\text{MTTF} = \frac{1}{\lambda}\left(\frac{1}{90} + \frac{1}{91} + \frac{1}{92} + \cdots + \frac{1}{100}\right) = 232 \text{ h}$$

(c) $$R(t) = \sum_{k=0}^{5} \frac{5^k e^{-5}}{k!} = 0.6157$$

$$\text{MTTF} = 123.11 \text{ h}$$

Switching in Redundant Systems

In redundant systems, it is necessary in many cases to isolate the failed components. This function of detection of failed components and isolation of these components from the network is performed by a *switching network*. In the analysis of the full-on, multiple-survivor redundant systems, the switching was considered to be perfect, that is, reliability of unity. But the switching network itself is an electronic circuit which has a reliability of less than unity. It is therefore of interest to analyze the reliability of redundant systems considering the effects of imperfect switching, that is, switch reliability that is less than unity.

In modeling the switching network, the question that arises is whether the switching network be considered as a component in series with the redundant system or as a component in parallel. From the definition of the series system it is evident that if any component fails, then the system fails. However, in this case, the switching network comes into play only at the time when one of the component fails, at which time the fault is detected (sensed) and isolated by the switching network.

It is clear that the switching network performs two distinct functions: (a) sensing or detecting the failed component; (b) isolating the failed component. Consider a situation where the sensing mechanism of the switching network has failed. The system as a whole will continue to operate successfully till one of the redundant units fail. On failure of a redundant unit, the switching network will be unable to detect the failed unit and therefore will not be able to isolate the failed unit. As a consequence, the system will cease to function satisfactorily. One should note that even though the sensing mechanism of the switching network had failed, the system did not immediately stop functioning. The system continued to operate satisfactorily till

the occurrence of the failure of one of the redundant units. Thus, one cannot say that the switching network is in series with the redundant system from the reliability point of view.

On the other hand, if one considers a situation wherein the switching network has failed in a manner that does not allow signal or power to flow through its switch, then the system would immediately fail. In such a case, the switching network can be considered as a component in series with the redundant system. Thus, the switching network is governed by two modes of failure that leads to two failure probabilities. They are as follows:

1. The *dormant mode failure* of the switching network wherein the sensing mechanism fails and the system continues to work till the failure of one of the components of the parallel redundant system. This failure probability of the switching network, denoted by q_{sd} and $p_{sd} = 1 - q_{sd}$, is the probability that the switching network does not fail in the dormant mode, that is, probability that the switching network performs sensing and isolation successfully.
2. The *immediate failure mode* of the switching network wherein this type of failure in the switching network leads to instantaneous failure of the system. This failure probability of the switching network, denoted by q_{si} and $p_{si} = 1 - q_{si}$, is the probability that the switching network does not cause immediate breakdown of the system.

Thus, the redundant system with imperfect switching can be represented as indicated in Figure 14.18. The switching network is represented as two parts: One which causes immediate failure of the system (shown in series with the redundant system) and other that causes the dormant mode failure of the system.

The switching network behavior can be illustrated with the following example. Consider the circuit shown in Figure 14.19. It is required that two power supplies PS1 and PS2 be paralleled to deliver some specific power to the load. The power supplies PS1 and PS2 are individually capable of handling the full load power. But on paralleling, the electrical stresses on each of the power supplies are reduced, thereby enhancing the reliability. The system is said to function satisfactorily when at least one power supply is working satisfactorily. When any one of the power supply fails, the output voltage is required to be sensed and the failed power supply should be isolated from the circuit. These functions of sensing the fault and isolating the failed unit are done by a switching network as shown. The switching network consists of two

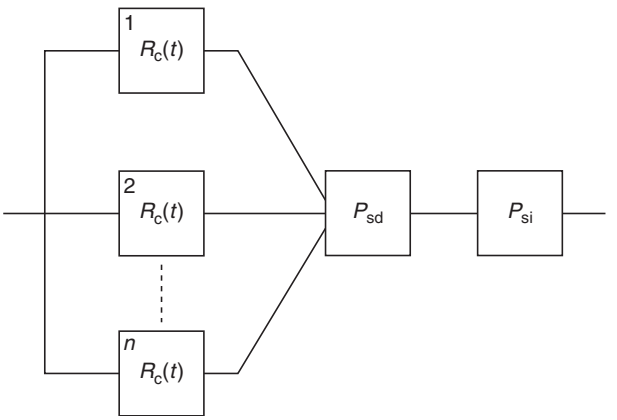


Figure 14.18 | Full-on, multiple-survivor redundant system with imperfect switching.

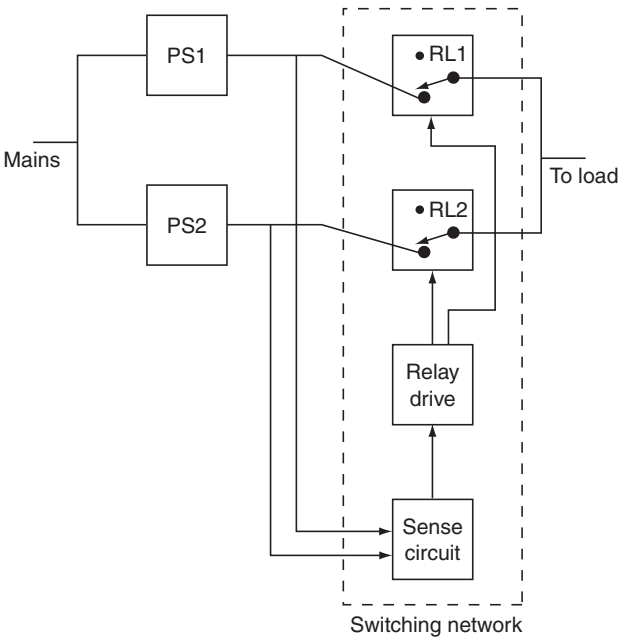


Figure 14.19 | Power supplies with relay network.

relays RL1 and RL2 through which power supplies PS1 and PS2, respectively, are connected to the load. The relays are controlled by a relay drive circuit. The control of the relay is based on the sense circuitry which senses the output voltage of the power supplies for fault condition.

The sensing circuitry of the switching network can fail. Then, in such a case depending on the type of relays RL1 and RL2 (normally closed or normally open), the system could continue to operate satisfactorily or not. The relay drive circuit may fail. Here again, depending on the relay types (i.e., normally closed or normally open), the system will continue to work or not.

The relay contacts may open circuit. In this case the system will fail immediately if both the relay contacts have failed. If only one relay contact has failed, then the system continues to work till the power supply corresponding to the working relay contact fails.

Thus it is evident that one type of failure of the switching network is the dormant mode failure and the other type of failure of the switching network causes immediate failure of the system. It is of interest to find the reliability of the full-on, multiple-survivor redundant system considering imperfect switching. Consider a three-component full-on, single-survivor redundant system as shown in Figure 14.20. To find the reliability of this system, one can first evaluate the reliability of the redundant system together with the dormant failure mechanism component of the switching network and then combine the resulting system with the immediate failure mechanism component of the switching network in a simple series network configuration.

Considering the system with only dormant failure mechanism part, the system works successfully

IF (no component fails) or
(1 component fails & the failure is successfully sensed and isolated) or
(2 components fail & both failures are successfully sensed and isolated)

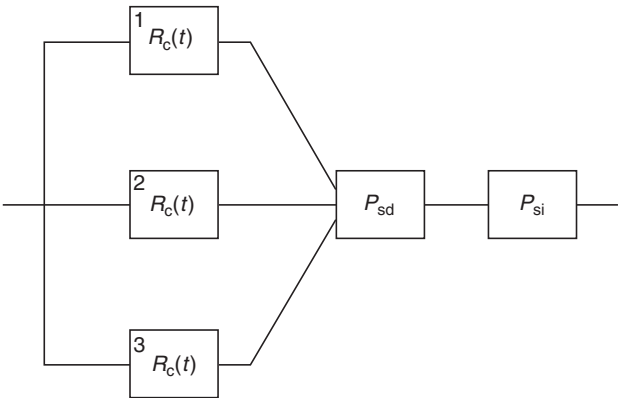


Figure 14.20 | Three-component, full-on, single survivor with imperfect sensing and switching.

Let X be a random variable that denotes the number of failures of components, then

$$R_s(t) = \mathrm{P}(X=0) + \mathrm{P}(X=1)\, p_{sd} + \mathrm{P}(X=2) p_{sd} p_{sd} \tag{14.52}$$

Using Eq. (14.46), Eq. (14.52) can be re-written as

$$R_s(t) = \sum_{k=0}^{2} 3C_k (1-R_c)^k (R_c)^{3-k} (p_{sd})^k \tag{14.53}$$

For full-on, m-survivor, n-component redundant system with imperfect switching, the reliability $R_s(t)$, as given by Eq. (14.53), can be generalized as

$$R_s(t) = \sum_{k=0}^{n-m} nC_k (1-R_c)^k (R_c)^{n-k} (p_{sd})^k \tag{14.54}$$

Considering the reliability of the immediate failure mechanism component of the switching network, p_{si}, the reliability $R(t)$ is given by

$$R(t) = \left[\sum_{k=0}^{n-m} nC_k (1-R_c)^k (R_c)^{n-k} (p_{sd})^k \right] p_{si} \tag{14.55}$$

Equation (14.54) defines the reliability of the full-on redundant system with dormant mode failure mechanism. One should note that if there are k units that have failed ($k < n - m$), then there must be k successful fault detection and isolation. Here p_{sd} provides practical boundaries for the reliability improvement due to the application of redundancy. A reasonable state-of-the-art value of $p_{sd} = 0.95$ is assumed to enable some comparisons to be made.

Referring to Figure 14.21, the reliabilities for full-on, single-survivor systems with and without perfect switching network are compared for various values of n. The light lines indicate the reliabilities for the case of perfect switching and the solid lines indicate the reliabilities for the case of imperfect switching. As can be seen from Figure 14.21, even for a $p_{sd} = 0.95$, a significant difference results between the reliability gained with perfect switching and that gained with imperfect switching. Further, one can note that as the number of redundant components increases, the difference is more pronounced.

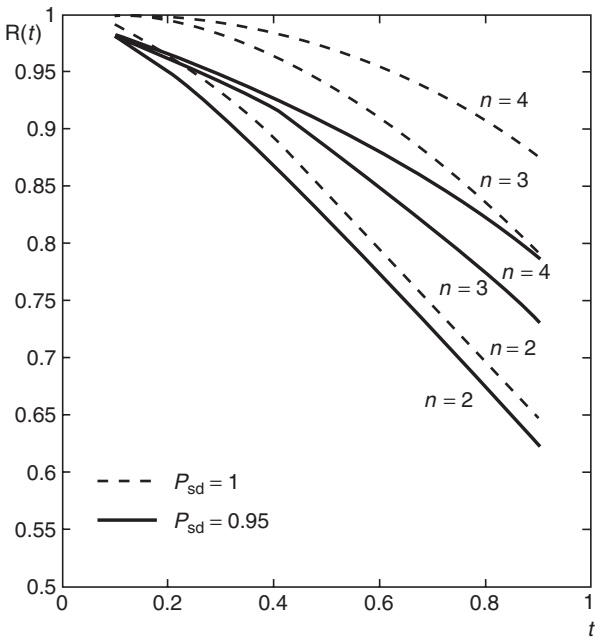


Figure 14.21 Full-on, *m*-survivor comparison with and without perfect sensing and switching network.

Mean Time to Failure

Considering the redundant system with only dormant mode failures of the switching network, the MTBF is obtained by using

$$\text{MTBF} = \int_0^\infty R(t)\mathrm{d}t$$

Substituting Eq. (14.54), one obtains

$$\text{MTBF} = \frac{1}{\lambda}\left(\frac{(p_{\text{sd}})^{n-m}}{m} + \frac{(p_{\text{sd}})^{n-m-1}}{m+1} + \quad + \frac{1}{n}\right) \tag{14.56}$$

One should note at this point that for $p_{\text{sd}} = 1$ (i.e., perfect switching), the MTBF of the system reduces to that of a full-on, multiple-survivor redundant system with perfect switching.

Problem 14.6

A system comprises two components in full-on, single-survivor redundant configuration. The failure rate of each component is 0.001 failures/hour. For a period of operation of 2 h, what is the

(a) reliability of a single component?
(b) reliability of the system with perfect switching?
(c) reliability of the system with $p_{sd} = 0.2$ and $p_{sd} = 0.5$?

Solution

(a) Let $R_c(t)$ be the reliability of each of the component. Then,

$$R_c(t) = e^{-0.001\times 2} = 0.998$$

(b) Let $R(t)$ be the reliability of the system. Then,

$$R(t) = 1 - (1 - R_c)^2 = 0.99999$$

(c) With $p_{sd} = 0.2$,

$$R(t) = \sum_{k=0}^{1} 2C_k (1-R_c)^k (R_c)^{2-k} (0.2)^k = 0.984$$

With $p_{sd} = 0.5$,

$$R(t) = \sum_{k=0}^{1} 2C_k (1-R_c)^k (R_c)^{2-k} (0.5)^k = 0.998$$

Therefore, it is evident that with $p_{sd} > 0.5$ the redundant system is more reliable than a single component. As a consequence it would make sense to use redundancy in this case only if the switching network has a reliability of greater than 0.5.

Problem 14.7

A system comprises two identical units which are connected in full-on, single-survivor redundant configuration. If the failure rate of one of the units is 0.001 failures/h, what should be the minimum reliability requirement for the switching network, with regard to dormant failures, for a period of 5 h of continuous operation, so that the redundant system is meaningful? (Assume that the switching network does not give rise to immediate failure of the system).

Solution

Let $R_c(t)$ be the reliability of each unit and p_{sd} the reliability of the switching network with regard to the dormant failures. For the redundant system to be meaningful, the reliability of the redundant system should be more than the reliability of the single system.

$$\sum_{k=0}^{1} 2C_k (1-R_c)^k (R_c)^{2-k} (p_{sd})^k > R_c$$

where $R_c = e^{-0.001 \times 5} = 0.995$.

Solving the above equation for p_{sd}, one obtains the condition that $p_{sd} > 0.5$ should be satisfied if the redundant system is to have a reliability greater than the reliability of the individual components.

Standby Redundancy

Another frequently used redundancy configuration is the standby redundancy which is shown in Figure 14.22. Let there be n components in a system. Out of these n components, m components are required to function for satisfactory operation of the system. The remaining $n - m$ components are kept in standby. These $n - m$ components are not fully energized as in the case of the full-on redundancy. Thus, in the standby mode of operation, the failure rate of the standby units can be assumed to be zero. One should note that Figure 14.22 is not a reliability block diagram but the functional topology of the system. But from the reliability point of view, the m-block unit in operation is a series network which is indicated in Figure 14.23.

When one of the m operating units fails, the failure of the unit is detected and the failed component is immediately replaced by one of the standby units. The system continues to perform its function until the store of standby units is exhausted. Then on the immediate next failure of one of the components, the system would fail. Thus the system can withstand upto $n - m$ failures and on the occurrence of $n - m + 1$ failures, the system fails.

In the analysis of the standby redundancy with multiple failure, the following points must be noted.

1. The occurrence of failure among the m operational units is a Poisson process, that is, at any given instant no two units will fail simultaneously.
2. There are m units functioning at any given time. If any one of the m functioning units fails, then the block of m-unit sub-system experiences a failure. Therefore, the m-unit sub-system forms a series network with an equivalent failure rate of $m\lambda$, where λ is the failure rate of the individual units.

If one of the m units fails, then the failed unit is replaced by a standby unit thus resurrecting the m-unit block immediately. The reliability of such an m-unit sub-system is dependent on the number of resurrections possible. As there are $n - m$ standby units, $n - m$ resurrections of the m-unit block are possible. Thus the system can withstand up to $n - m$ failures. For any failure more than $n - m$, there will be less than m functioning units and the system fails.

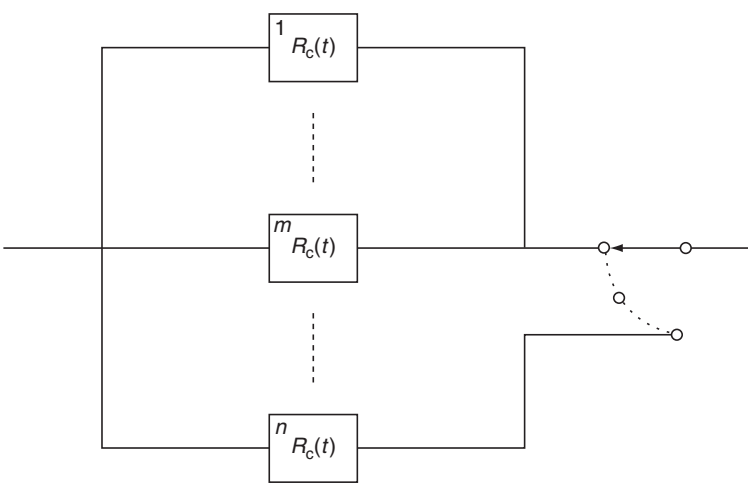


Figure 14.22 The m-survivor standby redundancy topology.

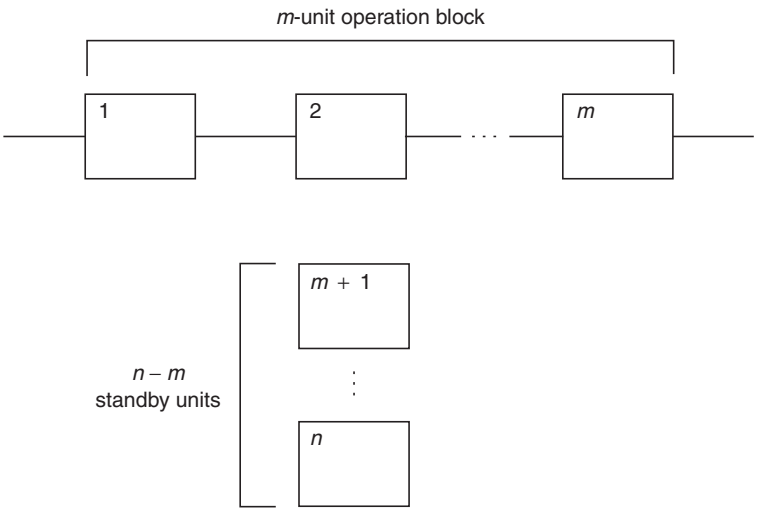


Figure 14.23 The *m*-survivor standby redundancy reliability block diagram.

The expected number of failures in time t is given by $m\lambda t$. Let X be the random variable that indicates the number of failures in a given time t. Then,

$$R(t) = \mathrm{P}(X \le n-m) = \sum_{k=0}^{n-m} \frac{(m\lambda t)^k e^{-m\lambda t}}{k!} \tag{14.57}$$

MTBF for Standby Redundancy Multiple-Survivor Configuration

The MTBF can be found by using

$$\mathrm{MTBF} = \int_0^\infty R(t)\mathrm{d}t$$

Substituting Eq. (14.57), one obtains

$$\mathrm{MTBF} = \frac{n-m+1}{m\lambda} \tag{14.58}$$

Alternatively the MTBF can also be obtained intuitively in the following manner.

$$\text{Mean time to first failure} = \frac{1}{m\lambda}$$

$$\text{Mean time to second failure} = \frac{1}{m\lambda} + \frac{1}{m\lambda} = \frac{2}{m\lambda}$$

$$\text{Mean time to third failure} = \frac{1}{m\lambda} + \frac{1}{m\lambda} + \frac{1}{m\lambda} = \frac{3}{m\lambda}$$

$$\text{Mean time to } n-m+1 \text{ failure} = \frac{n-m+1}{m\lambda}$$

This last equation is identical to Eq. (14.58).

Standby Redundancy with Imperfect Switching

Let p_{sd} be the probability that the switching network successfully senses a faulty unit and switches in a standby unit. By using a similar argument used for the case of full-on, multiple-survivor redundancy with imperfect switching, the reliability for the standby, multiple-survivor redundancy is given by

$$R(t) = \sum_{k=0}^{n-m}\left[\frac{(m\lambda t)^k e^{-m\lambda t}}{k!}(p_{sd})^k\right] \tag{14.59}$$

The MTBF for standby redundancy taking into account imperfect switching is given by

$$\text{MTBF} = \frac{1}{\lambda}(1 + p_{sd} + {p_{sd}}^2 + \cdots + {p_{sd}}^{n-m}) \tag{14.60}$$

It can be noted that for perfect switching when $p_{sd} = 1$, the MTBF reduces to Eq. (14.58).

14.4 Redundant Systems with Non-Identical Components

The preceding section has assumed that all the components of the standby system are identical. In a practical system, this is not necessarily true in all cases. The failure rates of the operating and standby components can be different owing to the nature of the components used. A common example is when a generator is used as the normal operating component of a DC supply and a battery is used as the standby component. In this case the failure rates could be very different. Further, in many cases, the failure rates of the standby components will not be zero. For example, in the case of an off-line UPS system, the battery–inverter system which is in standby till the mains power fails, will be generally in hot standby, that is, the battery–inverter system is energized but it operates under no-load. Once the mains power fails, the battery–inverter system will take on the full load. Thus, in such cases, the failure rate during standby is not zero but a finite value, which would of course be much lower than the failure rate when delivering full load power.

The approach used to model these systems is a direct and intuitive one. This approach is perfectly general and involves very little effort and computation time. In this method all events leading to system success are divided into mutually exclusive events. The expression for system reliability can then be derived by adding the reliability associated with each of these individual mutually exclusive events. This method is explained with the help of a few examples.

EXAMPLE 14.1

Consider the case of two non-identical components 1 and 2 forming the standby system in which component 2 is the standby component and cannot fail in the standby mode. It is of interest to find the reliability of the system till time t.

Let λ_1 be the failure rate of component 1 in the energized mode. Let λ_2 be the failure rate of component 2 in the energized mode. The failure rate of component 2 in the standby mode is zero.

The system operates successfully

IF (1 is good till t)

or (1 fails at $t_1 < t$) & (2 is good for $t - t_1$ period)

Let

$$R_1(t) = \text{P(1 is good till } t\text{)}$$

$$R_2(t) = \text{P[(1 fails at } t_1 < t\text{) \& (2 is good for } t - t_1 \text{ period)]}$$

Let $R(t)$ be the reliability of the system. As $R_1(t)$ and $R_2(t)$ are mutually exclusive,

$$R(t) = R_1(t) + R_2(t)$$

$$R_1(t) = e^{-\lambda_1 t}$$

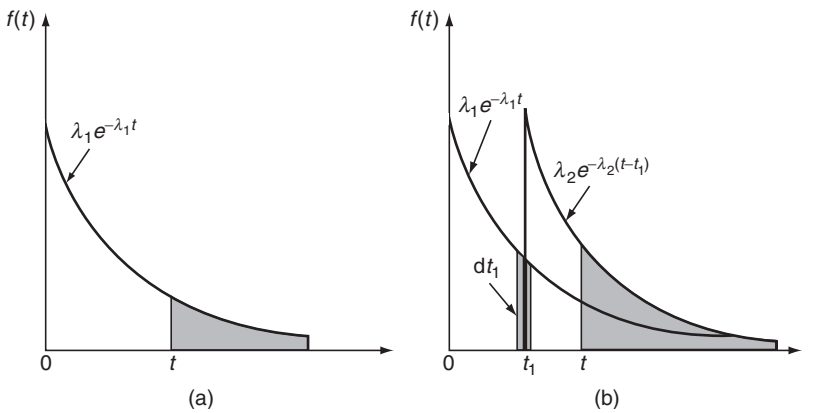


Figure 14.24 Graphical representation of reliability contributions: (a) Contribution of $R_1(t)$; (b) contribution of $R_2(t)$.

Referring to Figure 14.24, it is evident that the probability that component 2 is good for the period of time $(t - t_1)$ is the area shown shaded below the failure-time probability density function of component 2 after time t. The probability that component 1 fails at time t_1 is given by the area under the infinitesimal interval dt_1. However, one should note that the failure of component 1 at time t_1 can occur at any time between 0 and t. Therefore, the area below the failure-time probability density function of component 2 for every value of t_1 between 0 and t should be integrated. Thus,

$$R_2(t) = \int_{t_1=0}^{t} \lambda_1 e^{-\lambda_1 t} \cdot e^{-\lambda_2 (t - t_1)} \cdot dt_1$$

$$= \frac{\lambda_1}{\lambda_2 - \lambda_1} [e^{-\lambda_1 t} - e^{-\lambda_2 t}]$$

$$R(t) = e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1} [e^{-\lambda_1 t} - e^{-\lambda_2 t}]$$

Problem 14.8

Consider the situation as discussed in Example 14.1, but in this case the failure rate of the component 2 when in standby is λ_{2s}. Using similar logic, what is the reliability of the system?

Solution

The system operates successfully

IF (1 is good till t)

or (1 fails at $t_1 < t$) & (2 is good in standby till t_1) & (2 is good in energized mode for $t - t_1$ period)

$R_1(t)$ = P(1 is good till t)

$R_2(t)$ = P[(1 fails at $t_1 < t$) & (2 is good in standby till t_1) & (2 is good in energized mode for $t - t_1$ period)]

If $R(t)$ is the reliability of the system, then

$$R(t) = R_1(t) + R_2(t)$$

$$R_1(t) = e^{-\lambda_1 t}$$

$$R_2(t) = \int_{t_1=0}^{t} \lambda_1 e^{-\lambda_1 t} \cdot e^{-\lambda_{2s} t} e^{-\lambda_2 (t-t_1)} \cdot \mathrm{d}t_1$$

$$= \frac{\lambda_1}{\lambda_2 - \lambda_1 - \lambda_{2s}} [e^{-(\lambda_1 + \lambda_{2s})t} - e^{-\lambda_2 t}]$$

$$R(t) = e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1 - \lambda_{2s}} [e^{-(\lambda_1 + \lambda_{2s})t} - e^{-\lambda_2 t}]$$

Problem 14.9

Consider the system shown in Figure 14.25. Here the components 1 and 2 operate as a full-on, single-survivor redundant system and component 3 is used when both 1 and 2 have failed. For this system the following data is available:

(a) failure rate of component 1 when energized = λ_{1e}
(b) failure rate of component 2 when energized = λ_{2e}
(c) failure rate of component 3 when energized = λ_{3e}
(d) failure rate of component 3 when in standby = λ_{3s}
(e) failure rate of sensing device = λ_s
(f) failure rate of changeover device in position a = λ_{ea}
(g) failure rate of changeover device in position b = λ_{cb}
(h) probability of successful changeover = p_s

What is the reliability of the system?

Solution

As components 1 and 2 are operating in full-on, single-survivor redundant configuration, component 3 is switched into operation only after both components 1 and 2 have failed. To evaluate the reliability of the system, it is required to find the probability that the random variable T takes on values greater than t, that is, the system should not fail in the interval (0, t).

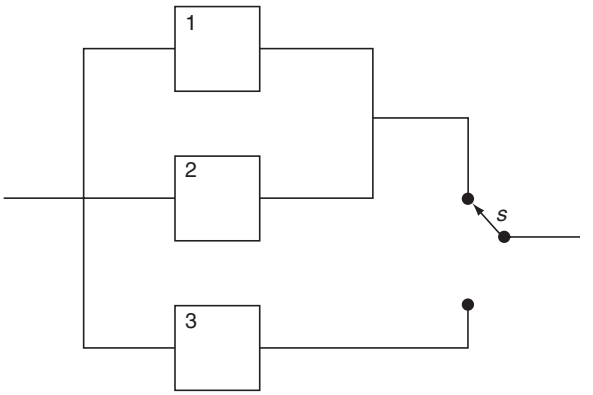


Figure 14.25 | System schematic for Problem 14.9.

The system operates successfully

IF (1 is good till t) & (2 is good till t) & (changeover device is good in position a till t)

or (1 is good till t) & (2 has failed before t) & (changeover device is good in position a till t)

or (1 has failed before t) & (2 is good till t) & (changeover device is good in position a till t)

or (1 fails at $t_1 < t$) & (2 fails at $t_2 > t_1$) & (3 is good in standby till t_2) & (3 is good for $t - t_2$ in energized mode) & (sensing is good till t_2) & (changeover device is good in position a till t_2) & (changeover device has successfully changed over) & (changeover device is good in position b for $t - t_2$)

or (2 fails at $t_1 < t$) & (1 fails at $t_2 > t_1$) & (3 is good in standby till t_2) & (3 is good for $t - t_2$ in energized mode) & (sensing is good till t_2) & (changeover device is good in position a till t_2) & (changeover device has successfully changed over) & (changeover device is good in position b for $t - t_2$)

In the above description of events $t_1 < t_2 < t$. Let

$$R_1(t) = e^{-\lambda_{1e}t} \cdot e^{-\lambda_{2e}t} \cdot e^{-\lambda_{ca}t}$$

$$R_2(t) = e^{-\lambda_{1e}t} \cdot (1 - e^{-\lambda_{2e}t}) \cdot e^{-\lambda_{ca}t}$$

$$R_3(t) = (1 - e^{-\lambda_{1e}t}) \cdot e^{-\lambda_{2e}t} \cdot e^{-\lambda_{ca}t}$$

$$R_4(t) = \int_{t_2=0}^{t} \left[\int_{t_1=0}^{t_2} \lambda_{1e} e^{-\lambda_{1e}t_1} \mathrm{d}t_1 \right] \cdot \lambda_{2e} e^{-\lambda_{2e}t_2} \cdot e^{-\lambda_{3s}t_2} \cdot e^{-\lambda_{3e}(t-t_2)} \cdot e^{-\lambda_s t_2} \cdot e^{-\lambda_{ca}t_2} \cdot p_s \cdot e^{-\lambda_{cb}t_2} \mathrm{d}t_2$$

$$R_5(t) = \int_{t_2=0}^{t} \left[\int_{t_1=0}^{t_2} \lambda_{2e} e^{-\lambda_{2e}t_1} \mathrm{d}t_1 \right] \cdot \lambda_{1e} e^{-\lambda_{1e}t_2} \cdot e^{-\lambda_{3s}t_2} \cdot e^{-\lambda_{3e}(t-t_2)} \cdot e^{-\lambda_s t_2} \cdot e^{-\lambda_{ca}t_2} \cdot p_s \cdot e^{-\lambda_{cb}t_2} \mathrm{d}t_2$$

If $R(t)$ is the reliability of the system, then

$$R(t) = R_1(t) + R_2(t) + R_3(t) + R_4(t) + R_5(t)$$

Problem 14.10

Consider the parallel connection of the two MOSFETs as shown in Figure 14.26. The required load current is 50 A. What is the reliability of the system if each MOSFET is rated

(a) for 30A?
(b) for 60A?

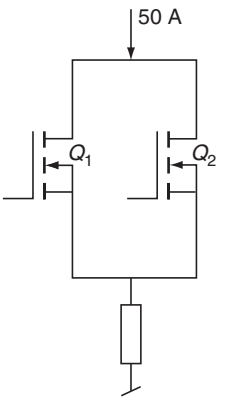


Figure 14.26 | Paralleling of MOSFETs.

Solution

When both the MOSFETs are operating, each of the MOSFET carries half the current (i.e., 25 A). Let λ_{25} and λ_{50} be the failure rates of Q_1 and Q_2 when carrying 25 A and 50 A currents, respectively. Let the portions of the failure rates associated with open-circuit failures of the transistors be λ_{o25} and λ_{o50}, respectively.

Case I (MOSFETs rated for 30 A each): Here both Q_1 and Q_2 should operate successfully for the system to operate. Therefore, this is the series network from the reliability point of view. Thus,

$$R(t) = e^{-\lambda_{25}t} \cdot e^{-\lambda_{25}t} = e^{-2\lambda_{25}t}$$

Case II (MOSFETs rated for 60 A each): The system operates successfully

IF (Q_1 is good till time $t - \lambda_{25}$) & (Q_2 is good till time $t - \lambda_{25}$)

or (Q_1 fails at $t_1 < t$ due to open-circuit failure $- \lambda_{o25}$) & (Q_2 is good till t_1 carrying 25 A $- \lambda_{25}$) & (Q_2 is good for $t - t_1$ carrying 50 A $- \lambda_{50}$)

or (Q_2 fails at $t_1 < t$ due to open-circuit failure $- \lambda_{o25}$) & (Q_1 is good till t_1 carrying 25 A $- \lambda_{25}$) & (Q_1 is good for $t - t_1$ carrying 50 A $- \lambda_{50}$)

Let

$R_1(t)$ = P[(Q_1 is good till time $t - \lambda_{25}$) & (Q_2 is good till time $t - \lambda_{25}$)]

$R_2(t)$ = P[(Q_1 fails at $t_1 < t$ due to open-circuit failure – λ_{o25}) & (Q_2 is good till t_1 carrying 25 A – λ_{25}) & (Q_2 is good for $t - t_1$ carrying 50 A – λ_{50})]

$R_3(t)$ = P[(Q_2 fails at $t_1 < t$ due to open-circuit failure – λ_{o25}) & (Q_1 is good till t_1 carrying 25 A – λ_{25}) & (Q_1 is good for $t - t_1$ carrying 50 A – λ_{50})]

As the events described above are mutually exclusive,

$$R(t) = R_1(t) + R_2(t) + R_3(t)$$

$$R_1(t) = e^{-2\lambda_{25}t} \quad \text{(using the third axiom of probability)}$$

$$R_2(t) = \int_{t_1=0}^{t} (\lambda_{o25} e^{-\lambda_{o25} t_1}) \cdot e^{-\lambda_{25} t_1} e^{-\lambda_{50}(t-t_1)} \cdot dt_1$$

$$= \frac{\lambda_{o25}}{\lambda_{50} - \lambda_{25} - \lambda_{o25}} [e^{-(\lambda_{25} + \lambda_{o25})t} - e^{-\lambda_{50} t}]$$

Similarly,

$$R_3(t) = \frac{\lambda_{o25}}{\lambda_{50} - \lambda_{25} - \lambda_{o25}} [e^{-(\lambda_{25} + \lambda_{o25})t} - e^{-\lambda_{50} t}]$$

Thus,

$$R(t) = e^{-2\lambda_{25}t} + \frac{2\lambda_{o25}}{\lambda_{50} - \lambda_{25} - \lambda_{o25}} [e^{-(\lambda_{25} + \lambda_{o25})t} - e^{-\lambda_{50} t}]$$

14.5 Design for Reliability

A primary attribute of any equipment is its ability to provide the required service under a specified range of constraints for a specified period of time, that is, the reliability of the equipment. The reliability of an equipment might be expected to be inversely proportional to its complexity, but this is not generally the case for equipments in large-scale production. A piece of equipment must reach a certain standard of reliability to gain acceptance. Hence with increasing complexity of equipment, the reliability of individual component must be correspondingly increased. In the field of electronics, new systems and new technologies are introduced before the engineering practices for assuring that their reliability has been fully evolved. In such cases, the contribution of electronic components to system reliability is more difficult to predict and where high reliability is required, the designer uses his best judgment in selecting components of assured quality and operating them under less stringent conditions.

Reliability, which is a measure of quality, is an essential element at each stage of the equipment manufacturing procedure, right from the initial specification of the equipment through the design and production stage, to the final delivery to the user. In designing a piece of equipment for improved reliability, the factors

to be considered are manifold and interacting. Some of these factors that affect the reliability of the equipment are the following:

1. **Design Process:** The methodology of designing equipment at the system level has a significant impact on the reliability of the equipment. In this section, a systematic approach to design of a given equipment will be illustrated using a case study.
2. **Inter-connection/Wiring:** One of the major factors that affect the reliability of the equipment is the manner in which the various modules of the system are inter-connected during the design process. One should give lot of consideration while selecting the type of connector or wires (conductors) for a given application.
3. **Effect of Circuit Design on Reliability:** The circuit must be designed taking into consideration the various component tolerances at all extremes of the operating conditions. The function of every component must be related to its capabilities as given in the published datasheet.
4. **Component Reliability:** The selection of appropriate component types is essential in equipments that are designed for reliability. As the reliability of components are reduced by operating them at ratings above those recommended, the components must be chosen such that they are operated well below the recommended ratings.
5. **Environmental Effects on Reliability:** Components must be provided with suitable layout, adequate heat sinking and protection against voltage transients and surges. The physical handling of the equipment and any special requirements of the equipment must be borne in mind, as well as the ambient conditions under which the equipment is to operate.

The main objective of the design process is to combine the various demands and constraints in an appropriate manner to result in the conception of the product. The various demands and constraints that are to be taken into account while designing a product/equipment will be the input to the design process. This input is called the *design specifications* or the *target specifications* that the final product should meet. The target specifications reflect the needs of the equipment user. These specifications are generally obtained by intensive market survey, user survey and survey of contemporary products.

The design process sets in after the target specifications are finalized. The general attitude to design of an electronic product is rather ad hoc. Once the specifications are obtained, most amateur designers would immediately start rigging up the relevant electronic circuits, fabricating the pcbs and integrating them to build the product. Such an approach would lead to solutions being *added to* rather than *integrated in* the design. For example, one may have ignored the target specifications and forgotten to incorporate a monitoring circuit or a protection circuit for a specific signal. This discrepancy may be noticed rather late in the development. As a consequence, the designer may add some extra circuits to correct the oversight. This would result in a poorly designed product from the reliability and the cost point of view. One should approach the design of the products in a more systematic manner wherein all the demands and constraints of the specifications are taken into account and inherently integrated within the design.

Further, one should note that there is no unique solution to achieve an intended function in electronic products. As there are many solutions, the question of which is the best solution arises. From the point of view of the product, the circuit configuration or module configuration that achieves the highest overall reliability to cost ratio would be the best choice. In view of the non-uniqueness in the approaches to design a specific system, this section will illustrate with a case study a sample design process keeping a measure of algorithmic style to the design process such that it can be extended to other applications.

Belt-Conveyor System: An Example Design Process

A belt-conveyor system is used to transfer material from one point to another. An induction motor is used to operate the belt-conveyor system as shown in Figure 14.27. It is required that the material should be transferred from one point to another at a user-defined fixed rate (kg/s). A product/equipment has to be developed that controls the speed of the induction motor in such a manner as to keep the feed rate constant at the user-defined value. The belt-conveyor system is in an environment that is not conducive for human presence. Therefore, the setting of the feed rate by the user has to be done at the remote location. A load cell is used as a transducer to measure the weight. A speed transducer that is mounted on the induction motor shaft is used to measure the conveyor speed. The following specifications are obtained in the context of this system.

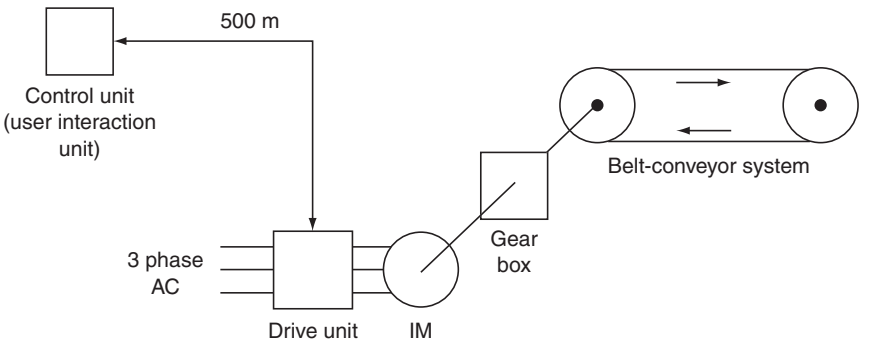


Figure 14.27 | Schematic of the belt-conveyor system.

1. Supply voltage: 3-phase, 400 V AC, 50 Hz, ±15%
2. Induction motor: 3-phase, 400 V, 4 pole, 5 HP machine
3. Speed range required: 10–90% of base speed
4. Distance between the control and drive: 500 m
5. Ambient temperature: 50°C
6. MTBF: 1 year

Protection and indication should be provided for over-voltage, over-heating, under-voltage and over-current.

Stimuli and Responses: To evolve a systematic methodology to the design of electronic equipments, the outer-to-inner approach is used. In this approach, the equipment is viewed first as a black box and gradually the various modules and sub-modules are segregated. The actual electronics design of the various sub-modules is the innermost stage of the outer-to-inner approach.

In this case, there are two sub-systems: (a) the control unit or the user interaction unit and (b) the drive unit. The control unit is where the user will interact with the system in terms of setting the feed rate, monitoring the faults, etc. Therefore, this unit is also called the user-interaction unit. The drive unit performs the actual speed control of the induction motor. There are two possible ways in which the two sub-systems can be located.

1. The control unit and the drive unit can be located away from the induction motor driven belt-conveyor system. In this case, the output of the drive unit which drives the induction motor is situated about 500 m from the induction motor. Power will be transmitted through the 500 m distance and as a consequence, heavy power cable will have to be used which will increase the cost of the system. Further, as the drive output is pulse-width modulated (PWM), transmission line effects will result in reflections and waveform distortions in addition to the power loss in the line.
2. The control unit can be placed in the control room which is 500 m away from the belt-conveyor system. The drive unit is placed close to the induction motor. In this case, there are no transmission line effects and further the line loss is negligible as only the DC control signals are transmitted over the 500 m distance. Hence this scheme is preferred.

The electronic portion of the system has been physically divided into two sub-systems: (a) the control unit and (b) the drive unit. It is now required to design these two electronic sub-systems such that the targeted specifications are met. The first stage of the design process aims at categorizing all the stimuli to the system and all the responses from the system. This stage is very crucial because all the following processes (interconnection, electronic design, etc.) will be affected by the outcome of this stage. Therefore, one must provide sufficient time to contemplate on the product so that all possible stimuli that the system may experience are considered. Likewise, one should account for all possible responses from the system.

Figure 14.28 shows the stimuli and the responses for both the control unit and the drive unit. Those stimuli and responses that are marked with a "*" are the critical stimuli and responses which correspond to the primary function of the system. Consider first the control unit.

Referring to Figure 14.28, the stimuli for the control unit are as follows:

1. **230 V AC:** This is the mains input which is essential for developing the control power for the entire electronics of the control unit.
2. **Feed Rate set:** This is the user-defined feed-rate setting.
3. **Weight Feedback:** A load cell is used to measure the weight on the belt-conveyor at a given instant of time.
4. **Speed from Tacho:** The speed of the induction motor, which is a direct measure of the speed with which the belt-conveyor is moving, is obtained from a tachogenerator that is mounted on the induction motor. The weight feedback and the speed feedback from the tacho are used to determine the feed rate of the belt-conveyor system.
5. **Power ON:** This stimulus is used to switch ON the control unit to enable interaction and control of the drive unit.
6. **Fault Reset:** When any fault occurs, the system is shut down. After maintenance and rectification of the fault, this stimulus is used to restart the system.
7. **Voltage Sense:** This is a stimulus that is obtained from the voltage sensors that are mounted in the drive unit. The DC-link voltage is measured and signal conditioned. This signal is used by the control unit to detect over- and under-voltage conditions.
8. **Current Sense:** This is a stimulus that is obtained from the current sensors that are mounted in the drive unit. The motor currents are measured and the conditioned signal is used as a stimulus for the control unit for the purpose of current control and over-current detection.
9. **Thermal Sense:** The temperature of the power-switching devices in the drive unit is sensed and signal conditioned. This signal is used by the control unit to detect over temperatures.

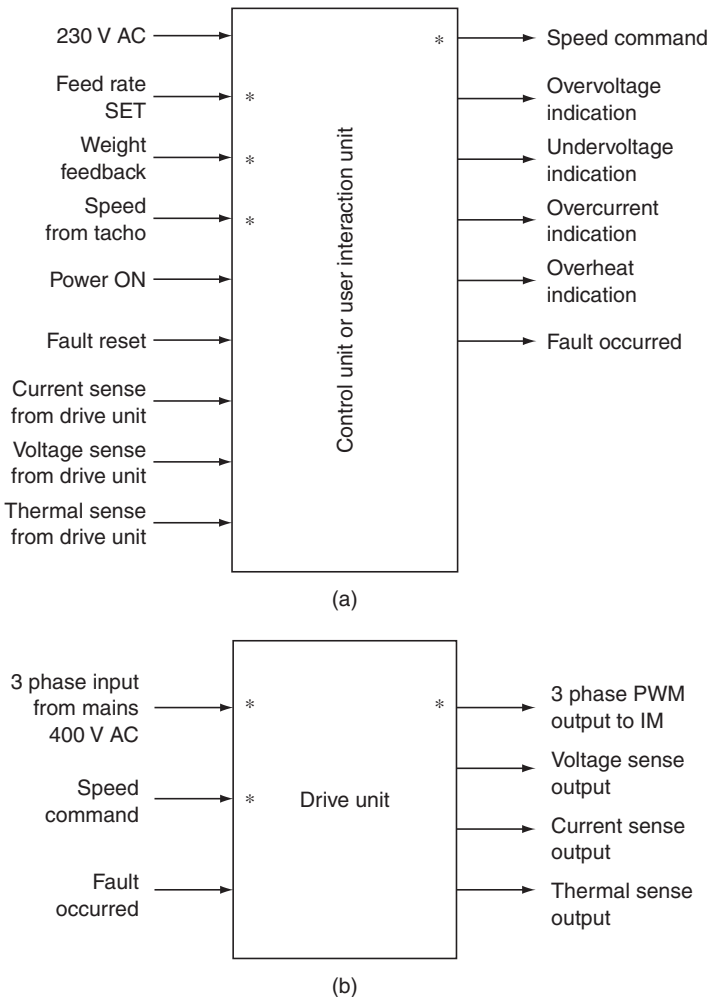


Figure 14.28 | Stimuli and response of the sub-systems.

The responses for the control unit are as follows:

1. **Speed Command:** The control unit must generate the command signal for the speed so that the drive unit can use this signal and appropriately control the speed of the induction motor.
2. **Over-Voltage Indication:** This is a response which is used to drive an LED that lights up when an over-voltage occurs.
3. **Under-Voltage Indication:** This is a response which is used to drive an LED that lights up when an under-voltage occurs.
4. **Over-Current Indication:** This is a response which is used to drive an LED that lights up when the load current (motor currents) exceed a certain value.

5. **Over-heat Indication:** This is a response which is used to drive an LED that lights up when the power devices in the drive unit exceed a certain temperature.
6. **Fault Occurred:** When any fault condition occurs (over- or under-voltage, over-current or over-temperature), then this response is asserted. This signal is used by the drive unit to shut down the system under fault conditions.

The stimuli for the drive unit are as follows:

1. **Three-Phase Input from Mains:** The power to drive the induction motor is drawn from the three-phase mains.
2. **Speed Command:** This stimulus is the response (or output) of the control unit. This is used to appropriately set the speed of the induction motor to achieve a feed rate equal to the SET feed rate.
3. **Fault Occurred:** This stimulus is also the response (or output) of the control unit. This is used to shut down the drive system when any fault occurs.

The responses of the drive unit are as follows:

1. **Three-Phase PWM Output:** This is the response (output) that is used to drive the induction motor.
2. **Current Sense:** This response is obtained by sensing the motor line currents by using a suitable current sensor and appropriately signal conditioning it.
3. **Thermal Sense:** This response is obtained by measuring the case temperature of the power-switching devices (or the heat-sink temperature on which the devices are mounted) with proper signal conditioning.

Functional Block Schematic: The functional block schematic of the system shows the essence of the system to be designed. Many of the functions such as indications, protections, etc., though they are important in a product, do not indicate the primary function. They are only subsidiary circuits that enhance the main function of the product. Therefore, to obtain a better understanding of the product, one must recognize those stimuli and responses that correspond to the primary function of the product. Thus referring to Figure 14.28, the stimuli and responses marked with "*" (called the primary stimuli and responses of the product) correspond to the primary function of the product. All other stimuli and responses are subsidiary.

From the primary stimuli and responses, the functional block schematic of the system can be developed. For the current case, the functional block schematic is shown in Figure 14.29. The functional segregation of the control unit and the drive unit is also shown in Figure 14.29. The weight from the load cell and the speed ω_m of the induction motor which is fed back is compared with the mechanical speed command from the output of the PI controller to obtain the speed error. The speed error is passed through another PI controller to obtain the slip speed ω_{sl}. The mechanical speed is added to the slip speed to obtain the synchronous speed ω_s. This ω_s becomes the speed command value that is the primary response of the control unit.

For the drive unit, the primary stimulus is the speed command value ω_s from the control unit. The desired stator voltage V_s for the corresponding ω_s is obtained from a lookup table. The desired V_s and ω_s are the inputs to the PWM modulator. The PWM modulator generates appropriate signals to switch the inverter switches. The three-phase mains is rectified and filtered to obtain the DC-link voltage for the inverter. The output of the inverter is the three-phase PWM output of the drive unit which is used to drive the induction motor.

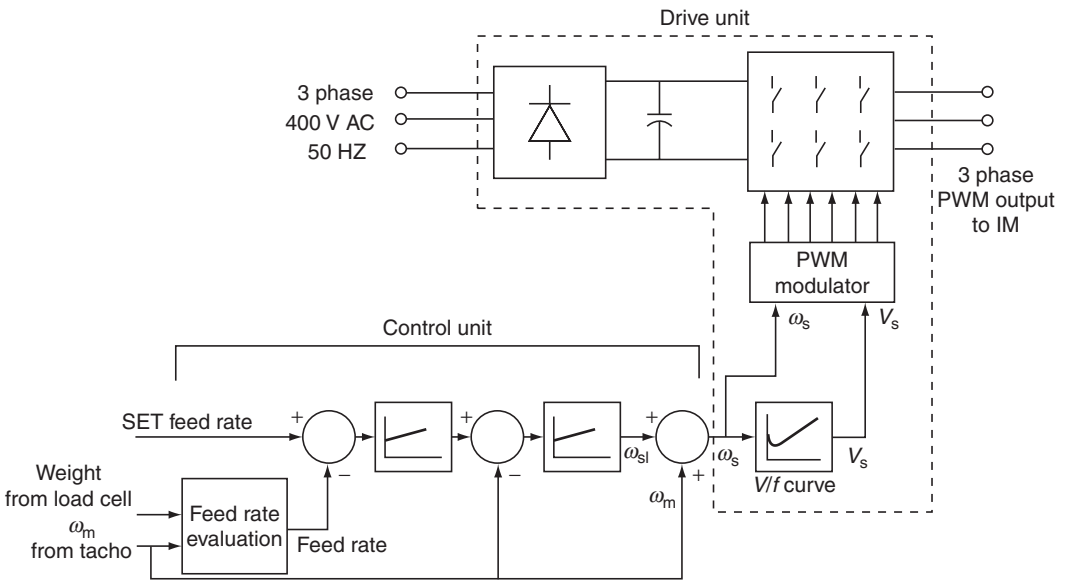


Figure 14.29 | Functional block schematic of the system.

Segregation into Modules: The system should now be broken up into smaller modules and sub-modules with clear stimuli and responses for each module as shown in Figures 14.30 and 14.31. At this point, it is also essential to indicate the number of wires or lines required for each stimulus or response. Note that if the number of wires for a particular stimulus or response is not indicated, then it is taken to be a single wire/line. It is also important to indicate the power supply requirements for each module at this stage. The segregation of modules for the control unit is shown in Figure 14.30 and the segregation of the modules for the drive unit is shown in Figure 14.31. Based on the overall target specification of the system, the designer should generate the specifications for each module at this stage.

The specifications for the modules of the control unit are as follows:

Power Supply Module

Stimuli:

Input: 230 V AC, 50 Hz
Range: 140 V AC to 270 V AC
Power ON switch: 230 V AC, 0.5 A rating with LED indication in switch

Responses:

Outputs: 5 V/2 A ± 5%
: +12 V/1 A ± 10%
: –12 V/1 A ± 10%

Controller Module

Stimuli:

SET feed rate: 0 to 10 V
Weight: 0 to 10 V (equivalent voltage output from load cell)
Speed: 0 to 10 V (equivalent voltage output from tacho)

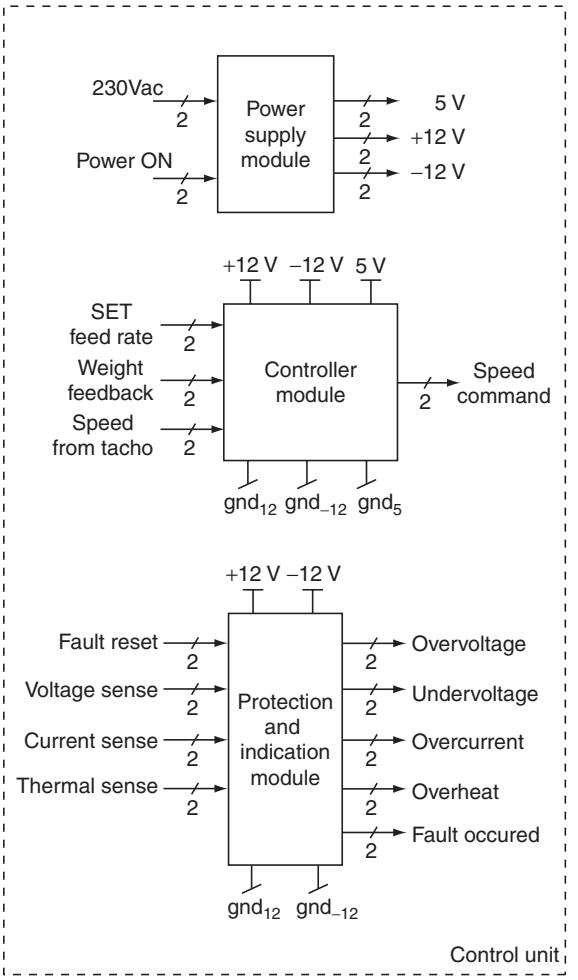


Figure 14.30 | Segregation into sub-modules for the control unit.

Response:

Speed command: 0 to 10 V

Protection and Indication Module

Stimuli:

Voltage sense: 0 to 2 V
Current sense: 0 to 2 V
Thermal sense: 0 to 2 V
Fault reset: Press to ON pushbutton switch

Responses:

Over-voltage: > 660 V DC-link
Under-voltage: < 500 V DC-link

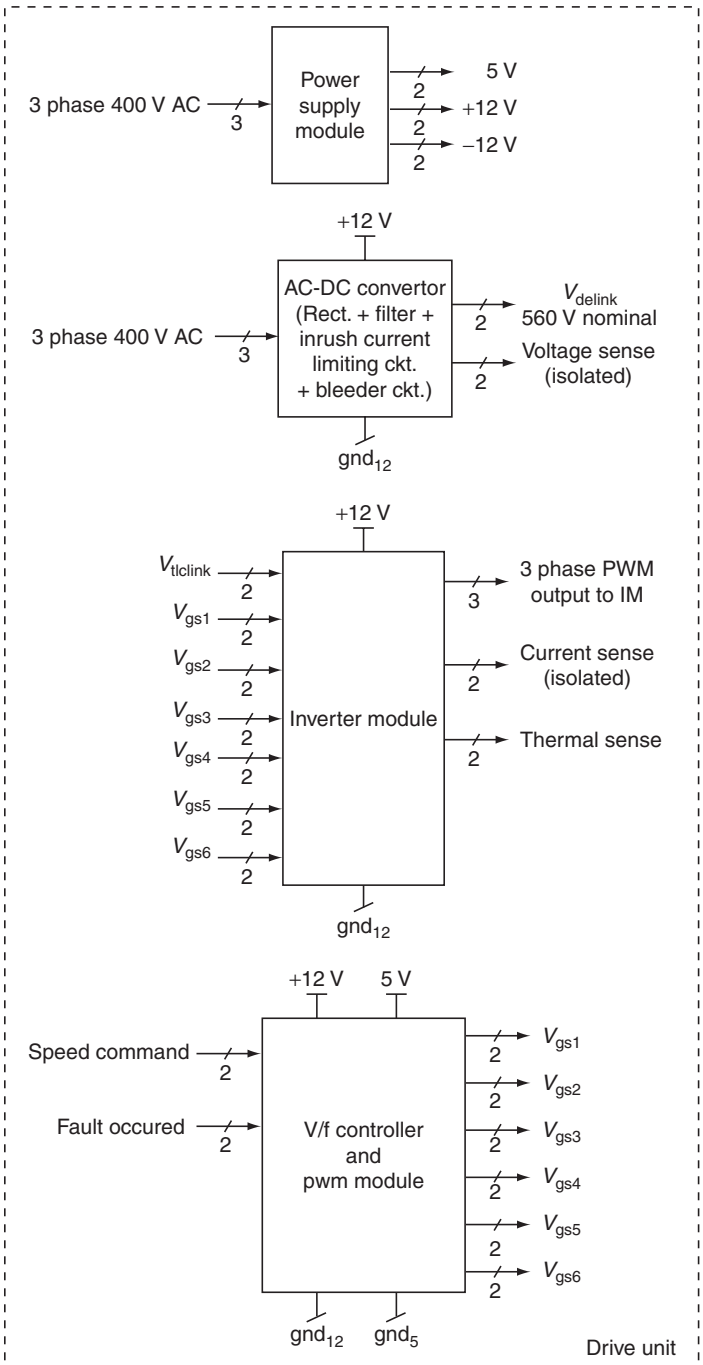


Figure 14.31 | Segregation into modules for drive unit.

Over-current: > 10 A line current of IM
Over-heat: > 75°C of inverter heat sink
Fault occurred: 0 V for not asserted
: 12 V for asserted (i.e., fault condition)

The specifications for the modules of the drive unit are as follows:

Power Supply Module
Stimuli:
Input: 3-phase, 400 V AC, 50 Hz, ±20%
Responses:
Outputs: 5 V/2 A ± 5%
: +12 V/1 A ± 10%
: –12 V/1 A ± 10%

AC–DC Converter Module
Stimuli:
Input: 3-phase, 400 V AC, 50 Hz, ±20%
Responses:
Outputs: 560 V DC nominal, ±20%
Output power: 5000 W
Voltage sense: 0 to 400 V AC (transformer isolated)

Inverter Module
Stimuli:
V_{dclink}: 560 V DC nominal, ±20%
$V_{\text{gs1...6}}$: 0 to 12 V
Responses:
Output: 3-phase PWM
Current sense: 0 to 10 A rms
Thermal sense: 0 to 75°C

***Vlf* Controller and PWM Modulator Module**
Stimuli:
Speed command: 0 to 10 V (represents 0 to 50 Hz)
Fault occurred: 0 V for no fault
: 12 V for occurrence of fault
Response:
$V_{\text{gs1}\cdots_6}$: gate drive
Signals: 0 to 12 V

14.6 Reliability Apportioning/Allocation

After the system has been segregated into various modules, it is now essential to allocate the desired reliabilities for each module or sub-module, such that the total specified system reliability is realized. It is therefore required to translate the reliability of the entire system into reliability requirements for the modules. This process is called reliability apportionment or allocation.

The allocation of system reliability involves solving the following basic inequality

$$f(R_1, R_2, \ldots, R_n) \geq R \tag{14.61}$$

where R_i is the allocated reliability for the ith module; R is the system reliability requirement; f is the functional relationship between the module and the system reliabilities.

For the case of a simple n-component series network system, if the reliabilities are considered for t h, then

$$R_1 \cdot R_2 \cdots\cdots R_n \geq R \tag{14.62}$$

where R_i is the allocated reliability of the ith module; R is the system reliability requirement.

For the case of an n-component, full-on, single-survivor redundant system, the reliability allocation inequality is given by

$$1-(1-R_1)\cdot(1-R_2)\cdots\cdots(1-R_n) \geq R \tag{14.63}$$

One should note that there are infinite solutions to the inequalities of Eqs. (14.62) and (14.63). However, it is only required that the allocation of the reliabilities be reasonable. The reliability allocation improves with experience. The reliability allocation gives a starting reliability value for the module design, which helps in choice of components, derating and design. After the design of the modules, the reliability prediction will give more accurate values of the module reliabilities. The reliability allocation and reliability prediction of the various modules and sub-modules are iterated till the desired system reliability is achieved.

If it is found that the reliability requirements of some individual modules are not met, the designer must use one or more of the following strategies:

1. Use components with lower hazard rate.
2. Simplify the design by using fewer component parts if it is possible without degrading the functional performance.
3. Apply component derating techniques to reduce the failure rates of the components (to be dealt later).
4. Use redundancy techniques.

Method 1: Equal Apportionment Technique

In the absence of definite information on the system other than the fact that the n modules are in series configuration, equal apportionment to each module would seem reasonable. In this case, the nth root of the system reliability requirement would be apportioned to each of the n modules. The main drawback in this approach is that the reliabilities are not allocated in accordance with the degree of difficulty associated with achieving these reliabilities for the various modules. For this method, the model is

$$R = R_1 \cdot R_2 \cdots\cdots R_n \tag{14.64}$$

Here $R_1 = R_2 = \ldots = R_n$. Thus,

$$R_i = R^{1/n} \quad \text{for } i = 1, 2, \ldots, n \tag{14.65}$$

where R is the required system reliability obtained from the target specifications; R_i is the reliability requirement apportioned to module i.

Method 2: AGREE Apportionment Technique

AGREE (Advisory Group on Reliability of Electronic Equipment) has outlined a method of apportionment for electronic equipments. This technique takes into consideration the complexity and importance of each

sub-system. It assumes a series of k sub-systems. The apportioned reliability goal is expressed in terms of MTBF. The minimum acceptable mean life of the ith sub-system is defined as:

$$\text{MTBF}_i = \frac{Nw_i t_i}{n_i[-\ln R(t)]} \tag{14.66}$$

and the corresponding ith sub-system reliability requirement is given as

$$R_i(t_i) = e^{-t_i/\text{MTBF}_i} \tag{14.67}$$

where $i = 1, 2, \ldots, k$; t is the required time for which the system should function (mission time); t_i the required mission time for the ith sub-system; w_i the importance factor that is expressed as the probability that the failure of the ith sub-system will result in the system failure, that is, P(system fails/ith sub-system fails); n_i the number of modules in the ith sub-system; N the total number of modules in the system (modules in all sub-systems put together); $R(t)$ the required reliability of the system for the system mission time; $R_i(t_i)$ the reliability apportioned to the ith sub-system for its mission time; MTBF_i the apportioned MTTF of the ith sub-system.

Method 3: ARINC Apportionment Technique

This method assumes series modules such that any module failure causes system failure and the module mission time is the same as the system mission time. This apportionment technique requires expression of reliability requirements in terms of failure rate.

The following steps are used in this technique of reliability apportionment:

1. The objective is to choose the failure rates of the modules such that

$$\sum_{i=1}^{n} \lambda_i \le \lambda$$

where λ_i is the failure rate allocated to the ith module and λ is the required system failure rate.

2. Determine the module failure rates (λ_{ie}) from past experience or from modules designed previously that has similar complexity.

3. Assign a weighting factor (w_i) to each module according to the failure rates determined in (2). Thus,

$$w_i = \frac{\lambda_{ie}}{\sum_{\text{all } i} \lambda_{ie}}$$

4. Allocate the module failure rate as

$$\lambda_i = w_i \lambda$$

For the belt-conveyor system being discussed, there are two sub-systems: the control unit and the drive unit. The control unit consists of three modules and the drive unit consists of four modules. It is now required to allocate reliabilities for these modules based on the apportionment techniques discussed, so that there is a starting reliability specification for the design of the electronic circuits for the various modules. For this case, method 3 can be used as there is some experience with products of similar nature and complexity.

The required MTBF that is specified is 1 year. Thus, the system failure rate should be less than 1 failure/year for 24 h a day operation. The various modules of the control unit and the drive unit are connected in series from the reliability point of view.

The reliability allocation table (Table 14.1) is based on the reliability apportionment method 3. All the symbols and relationships have the same meanings as discussed in method 3. It is found that the total failure rates of all the modules put together is equal to 0.75 failure/year, which is less than the required failure rate of 1 failure/year for the system. The allocated failure rates for each individual module is given in the last column, which is the product of the weighting factor w_i and the required failure rate of the system, that is, 1 failure/year in this case. Thus, the starting specifications for the failure rates of the various modules are as given in the allocated failure rates given in the last column of Table 14.1. Now the electronics for the various modules will be designed in a manner such that the allocated failure rates are met.

Table 14.1 | Failure rate allocation for the various modules of the belt-conveyor system

Sub-system	Modules	Failure rates from past experience (λ_{ie} failures/year)	Weighting factor (w_i)	Allocated failure rates (λ_i failures/year)
Control unit	Power supply	0.1	0.133	0.133
	Controller	0.1	0.133	0.133
	Protection and indication	0.05	0.066	0.066
Drive unit	Power supply	0.1	0.133	0.133
	AC–DC converter	0.15	0.2	0.2
	Inverter	0.2	0.266	0.266
	V/*f* controller and PWM modulator	0.05	0.066	0.066
Total		0.75	1	1

14.7 Inter-Connection and Wiring

In the previous sections, the modules were appropriately segregated and the reliability allocation for the individual modules was performed. The next step in the design process is to clearly specify the nature of the inter-connections between the various modules. In this context, it is first necessary to segregate the modules into printed circuit board (PCB) and non-PCB sub-modules. This exercise is necessary as it has a significant bearing on the nature of the inter-connection and wiring. The segregation of the various modules into PCB and non-PCB modules for the control unit is given in Table 14.2. This segregation into PCB or non-PCB sub-modules requires some experience and also some insight into the product as a whole. Like in the case of reliability allocation, this exercise of segregation of the modules into PCB and non-PCB sub-modules will improve with experience.

After segregating the various modules into the PCB and the non-PCB sub-modules, the various connectors and terminations should be labeled as indicated in Figures 14.32 and 14.33. Any consistent naming convention can be followed for the inter-connections. Here, the first letter J is reserved for PCB connectors and the first letter T is reserved for non-PCB terminations. This is followed by one or two letters to identify the module or sub-module and then followed by a letter count to indicate the number of connectors for a

Table 14.2 | Segregation of modules into PCB and non-PCB sub-modules

Module	Module segregation into sub-modules	Module/sub-module type	Remarks
Control Unit			
Power supply	–	PCB	–
Controller	–	PCB	–
Protection and indication	–	PCB	–
Drive Unit			
Power supply	–	–	–
AC–DC converter	Rectifier	Non-PCB	Heat-sink assembly
	Filter capacitors		–
	Inrush + Bleeder circuits		–
	Driver circuit for inrush and bleeder switches	PCB	–
Inverter	Inverter switches	–	Heat-sink assembly
	Current and thermal sense circuit	PCB	–
V/f controller and PWM modulator		PCB	–

given sub-module. The connectors and the associated signals on every pin of the connectors have to be decided. The connector symbolic names corresponding to the various modules are indicated in Figure 14.32 for the control unit and Figure 14.33 for the drive unit.

It can be noted from the Figures 14.32 and 14.33 that the first letter for the connector nomenclature is "J" for PCB-mountable connectors and "*T*" for non-PCB mountable terminations.

Connector nomenclature is followed by identification of every signal line on the connectors. It should be understood that the identification of the various signals with the respective connector pins will in general iterate with the PCB layout and module layouts. Referring to Figures 14.30 and 14.31, it can be noted that many of the stimuli and responses have two lines (signal and ground). Therefore, a "#" is used to indicate one of the signal lines and a "*" is used to indicate the other signal line. The last column of Tables 14.3 and 14.4 indicates which of the signal lines on the connector are connected to specific control power supply grounds.

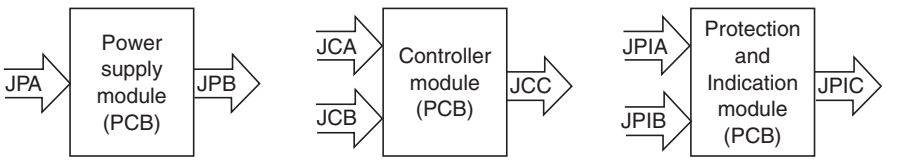


Figure 14.32 | Connector symbolic names for the modules of the control unit.

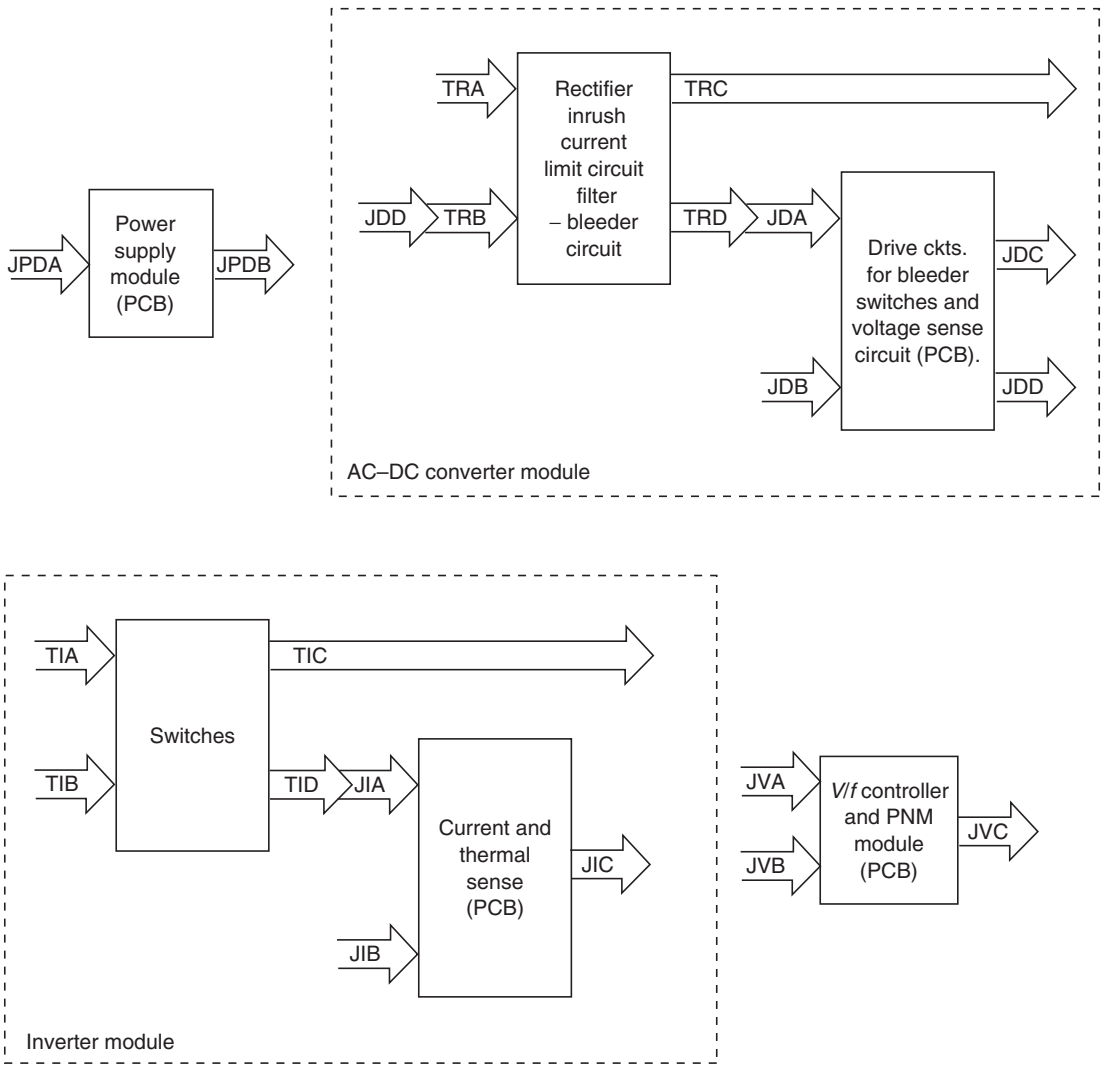


Figure 14.33 | Connector symbolic names for the modules of the drive unit.

The identification of the various signals of the various modules to the pins of the various connectors for the control unit is given in Table 14.3.

The identification of the various signals of the various modules with the pins of the connectors for the drive unit is given in Table 14.4.

Thus, the belt-conveyor system has been divided into sub-systems, that is, the control unit and the drive unit. All possible stimuli and the responses for the system are identified based on the target specifications. This is followed by segregation of the sub-systems into modules. The functional specifications for each module are identified. The reliability apportionment for every module of the system is performed. This is followed by the segregation of the various modules into PCB and non-PCB sub-modules. The identification of the various

Table 14.3 | Signal–connector pin details for modules of control unit

Module	Connector name	Pin no.	Signals		Control power supply
			Stimuli	Responses	
Power supply	JPA	1	230 V – L		
		2	230 V – N		
		3	Power on#		
		4	Power on*		
	JPB	1		5 V	
		2		gnd_5	
		3		12 V	
		4		gnd_{12}	
		5		–12 V	
		6		gnd_{-12}	
Controller	JCA	1	Set feed rate#		
		2	Set feed rate*		gnd_{12}
		3	Weight#		
		4	Weight*		gnd_{12}
		5	Speed#		
		6	Speed*		gnd_{12}
	JCB	1			12 V
		2			gnd_{12}
		3			–12 V
		4			gnd_{-12}
		5			5 V
		6			gnd_5
	JCC	1		Speed com#	
		2		Speed com*	gnd_{12}
Protection and indication	JPIA	1	Fault reset#		
		2	Fault reset*		gnd_{12}
		3	Voltage sense#		
		4	Voltage sense*		gnd_{12}
		5	Current sense#		
		6	Current sense*		gnd_{12}
		7	Thermal sense#		
		8	Thermal sense*		gnd_{12}
	JPIB	1			12 V
		2			gnd_{12}
		3			–12 V
		4			gnd_{-12}

(*Continued*)

Table 14.3 | Continued

Module	Connector name	Pin no.	Signals		Control power supply
			Stimuli	Responses	
	JPIC	1		Over-voltage	
		2		Under-voltage	
		3		Over-current	
		4		Over-heat#	
		5		Over-heat*	gnd_{12}
		6		Fault occurred#	
		7		Fault occurred*	gnd_{12}

Note: "#" is used to indicate one of the signal lines and "*" is used to indicate the other signal line.

Table 14.4 | Signal–connector pins identification for the modules of drive unit

Module	Connector name	Pin no.	Signals		Control power supply
			Stimuli	Responses	
Power supply	JPDA	1	400 V – R		
		2	400 V – Y		
		3	400 V – B		
	JPDB	1		5 V	
		2		gnd_5	
		3		12 V	
		4		gnd_{12}	
		5		–12 V	
		6		gnd_{-12}	
AC–DC converter	TRA	1	400 V – R		
		2	400 V – Y		
		3	400 V – B		
	TRB	1	Bleeder – V_{gs}#		
		2	Bleeder – V_{gs}*		gnd_{12}
	TRC	1		V_{dclink}#	
		2		V_{dclink}*	
	TRD	1		V_{dclink} – sense#	
		2		V_{dclink} – sense*	

(*Continued*)

Table 14.4 | Continued

Module	Connector name	Pin no.	Signals		Control power supply
			Stimuli	Responses	
	JDA	1	V_{dclink} – sense#		
		2	V_{dclink} – sense*		
	JDB	1			12 V
		2			gnd_{12}
	JDC	1		Voltage sense#	
		2		Voltage sense*	gnd_{12}
	JDD	1		Bleeder – V_{gs}#	
		2		Bleeder – V_{gs}*	gnd_{12}
Inverter	TIA	1	V_{dclink}#		
		2	V_{dclink}*		
	TIB	1	V_{gs1}#		
		2	V_{gs1}*		
		11	V_{gs6}#		
		12	V_{gs6}*		
	TIC	1		3-phase PWM – R	
		2		3-phase PWM – Y	
		3		3-phase PWM – B	
	TID	1		Line current#	
		2		Line current*	
		3		Heat sense#	
		4		Heat sense*	
	JIA	1	Line current #		
		2	Line current*		gnd_{12}
		3	Heat sense#		
		4	Heat sense*		gnd_{12}
	JIB	1			12 V
		2			gnd_{12}
	JIC	1		Current sense#	
		2		Current sense*	gnd_{12}
		3		Thermal sense#	
		4		Thermal sense*	gnd_{12}

(*Continued*)

Table 14.4 | Continued

Module	Connector name	Pin no.	Signals		Control power supply
			Stimuli	Responses	
V/f controller and PWM modulator	JVA	1	Speed com#		
		2	Speed com#		gnd_{12}
		3	Fault occurred#		
		4	Fault occurred*		gnd_{12}
	JVB	1			12 V
		2			Gnd_{12}
		3			5 V
		4			gnd_5
	JVC	1		V_{gs1}#	
		2		V_{gs1}*	gnd_{12}
		11		V_{gs6}#	
		12		V_{gs6}*	gnd_{12}

Note: "#" is used to indicate one of the signal lines and "*" is used to indicate the other signal line.

signals with the various connector pins for all the modules is performed so that all the stimuli, responses and the control power supply lines are identified and accounted. Now the electronics design for each of the modules can be performed to achieve the desired performance specification on the responses from the given stimuli. While designing the electronics for the various modules, one should take care that the allocated reliabilities for the individual modules are clearly met, such that the overall system reliability requirement is satisfied.

The design process that is discussed for the belt-conveyor system is a general procedure which can be adopted for any product design. One should note that the number of modules and the segregation of the modules are not unique for any product. However, the approach to design the product, irrespective of the non-uniqueness of the segregation of the modules, will be the same. Thus, one will be able to account for all the stimuli and responses indicated in the target specifications and as a consequence the designs will be integrated right from the beginning leading to faster design turn-around time.

14.8 General Notes on Designing for Reliability

While designing electronic circuits there are some precautions that one must take with regard to some key aspects in electronic design that will enhance the reliability of the electronic circuits. The following are some issues that are generally not addressed in the traditional functional design of the electronic circuits as most of these aspects are part of system integration.

Connectors

In any product, connections present a major problem from the point of view of reliability. Larger number of connectors implies lesser reliability. However, with more connectors, it becomes easier to maintain the product. Thus, connectors present a classic case of trade off between reliability and maintainability.

Table 14.5 | Relative failure rates for various connections

Connection	Index
Wrapped joint	1
Welded connection	3
Machine soldered joint (wave soldering)	7
Crimped joint	8
Hand soldered joint	10
Edge connectors (per pin)	30

Table 14.5 presents the various types of connections ranked in the order of reliability starting with the most reliable. An approximate idea of the relative failure rate is indicated by the index. The most reliable has an index of unity.

In this context, it is important to realize that the inter-connection stage of the design process must be iterated to optimize the number of connections. Apart from directly affecting the reliability of the product because of the inherent failure rates of the connectors, the connectors also affect the failure rate of product indirectly by adversely affecting the functioning of the circuits (especially high-frequency circuits) in the following ways:

1. Mutual inductance effect which causes crosstalk between signals.
2. Series inductance effect which slows down the signal propagation and creates electromagnetic interference (EMI).
3. Parasitic capacitance effect which slows down signal propagation.

Mutual Inductance Effect: Crosstalk

The mutual inductance effect between signal lines is illustrated in Figure 14.34. The current leaving gate A returns to its ground through the ground return path of the connector. This results in the loop area A. The current leaving gate B returns to its ground through the ground return path of the connector resulting in the loop area B. As the loop areas A and B overlap, magnetic fields from signal A induce electric voltages in signal path B and vice-versa.

The amount of crosstalk between any signal pins can be estimated if the following quantities are known:

1. The mutual inductance between the two loops.
2. The maximum rate of change of the source signal di/dt.
3. The impedance of the receiving network and whether it is source or end terminated.

The crosstalk between directly overlapping loops like the loop A and loop B is the worst-case crosstalk. The contributions to total magnetic flux in loop B come from two places. The first contribution is from currents flowing out of gate A along its signal wire. The second contribution is from returning signal currents flowing in the ground wire. The mutual inductance is given by

$$L_{A,B} = 5.08H \ln\left(\frac{c}{a}\right) + 5.08H \ln\left(\frac{b}{d/2}\right) \tag{14.68}$$

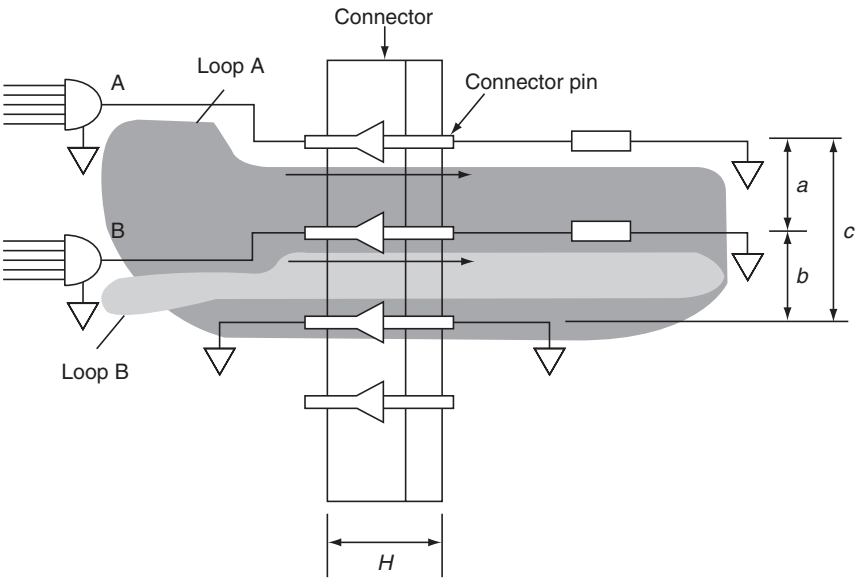


Figure 14.34 | Mutual inductive coupling effect in connectors.

where a is the distance of signal A to signal B (inches); b the distance of signal B to ground pin (inches); c the distance of signal A to ground pin (inches); d the diameter of connector pin (inches); H the pin length in the connector (inches); $L_{A,B}$ the mutual inductance between loops A and B (nH).

Equation (14.68) determines accurately enough whether or not connector crosstalk performance will be a significant issue. To estimate the amount of crosstalk, the maximum di/dt for the signals in question is required. This value can be measured using a high-bandwidth oscilloscope. The amount of crosstalk for drivers that are close to the connector is given by

$$\text{Crosstalk} = L_{A,B}\frac{di}{dt} \tag{14.69}$$

It is evident from Eq. (14.69) that slowing down the rise time of the driving signal reduces the crosstalk levels. The driving signal rise time can be reduced with a capacitor on the source side of the connector as shown in Figure 14.35. It does not flow through the connector. Therefore, the signal currents through the connector pins will have lower rise times which will drastically reduce the crosstalk between signals. Use of a resistor or an inductive bead as shown in Figure 14.35, improves the filtering and makes it more effective in reducing the crosstalk.

With respect to the crosstalk problems, which are predominant in high-speed digital circuits, the following points will aid in their solution:

1. Changing the pattern of the ground connections as indicated in Figure 14.36, one can increase or decrease the mutual inductance between specific lines. If the ground pin is moved further away from the signal lines A and B, the dimensions b and c will both increase. As a consequence, from Eq. (14.68) it becomes evident that the mutual inductance $L_{A,B}$ will increase. Conversely, moving the ground closer to the signals A and B will decrease their mutual inductance. The change in the inductance is proportional to the logarithm of the distance.

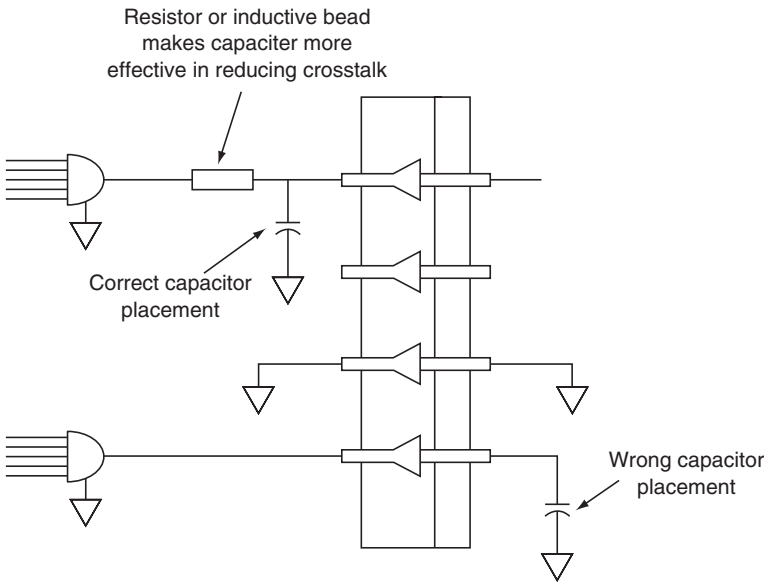


Figure 14.35 | Slowing the rise time of the signal by filtering.

2. Adding extra grounds will have a stronger effect. By adding more grounds, the return currents will divide among the various ground connectors thereby reducing the mutual inductance drastically. If a ground pin is placed above the signal line A besides the already existing ground pin below the signal line B as shown in Figure 14.36, the mutual inductance would drop almost by a factor of two. This is because the return currents will divide between the two grounds and reduce the mutual inductance. Further, the currents returning through the ground pin above signal A (loop A_1 and loop B_1) will be flowing in anti-clockwise direction and the currents returning through the ground pin below signal B (loop A_2 and loop B_2) will be flowing in clockwise direction. This would result in some cancellation of the fluxes and further reduce the mutual inductance. However, adding more ground pins above and below the A and B signal lines will not reduce the mutual inductance significantly any further.

Interposing ground pins between signal lines A and B as shown in Figure 14.37 makes a much bigger difference than adding grounds outside the signal lines A and B. If n ground pins are added between A and B as shown in Figure 14.37, then the coupling reduces drastically with n. This is given as follows:

The coupling is proportional to $1/(n^2 + 1)$. Noise coupled onto any given wire is contributed by each of the other wires in the connector. By reducing the number of signals in a connector, the aggregate crosstalk can be reduced. Alternately, partitioning a connector into several weakly interacting signal groups by placing intervening grounds between each group accomplishes the same thing. Adding extra grounds at the end of a connector does almost nothing to reduce the crosstalk. Large ground lugs at the end of a connector also will do nothing to reduce crosstalk.

Series Inductance Effect: EMI

Electromagnetic interference (EMI) emanates from signal currents flowing in large loops. This results in malfunction of the system, thereby affecting reliability. Referring to Figure 14.38, card A sources a 64-bit bus

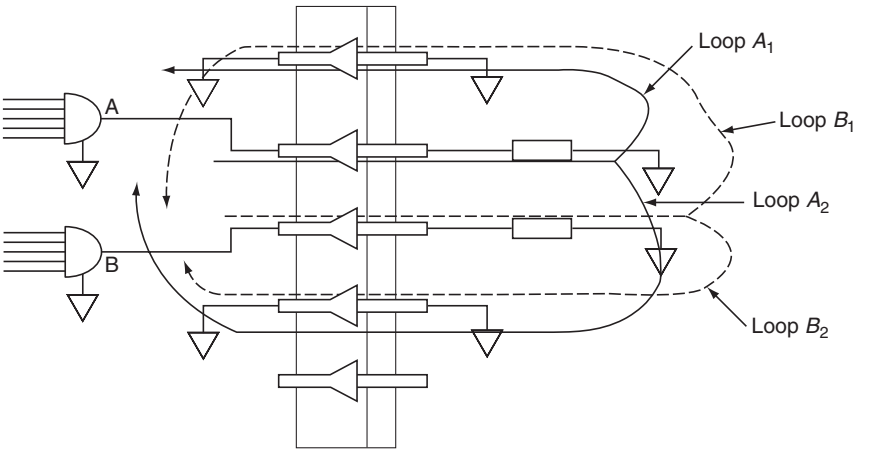


Figure 14.36 | Adding ground pins on either side of the signals.

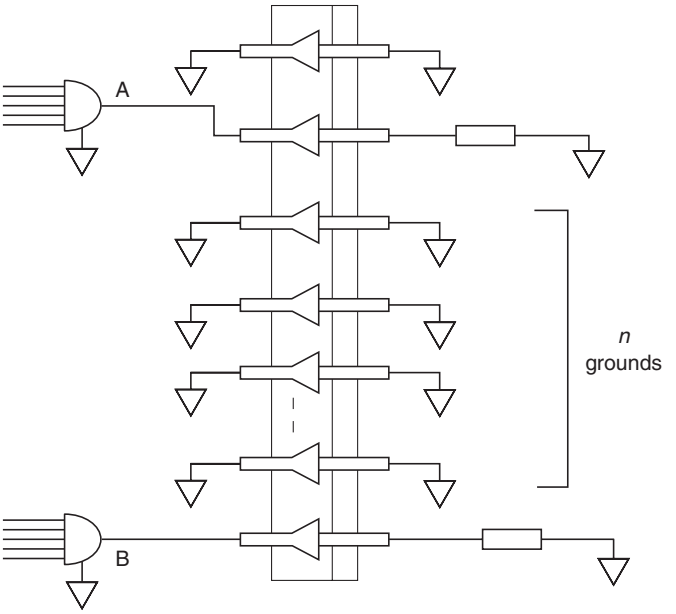


Figure 14.37 | Interposing ground pins between signal pins.

through connector 1 onto another card B. Signal return currents from these 64 lines flows from card B to card A mostly through the ground pins in connector 1. However, small fractions of the returning signal current flows back to card A through a different path. It may take a path through one of the ground pins of connector 2 or through some other remote path which may flow through many other cards. It is this small fraction of the return currents which takes these longer return paths that cause the major problem due to

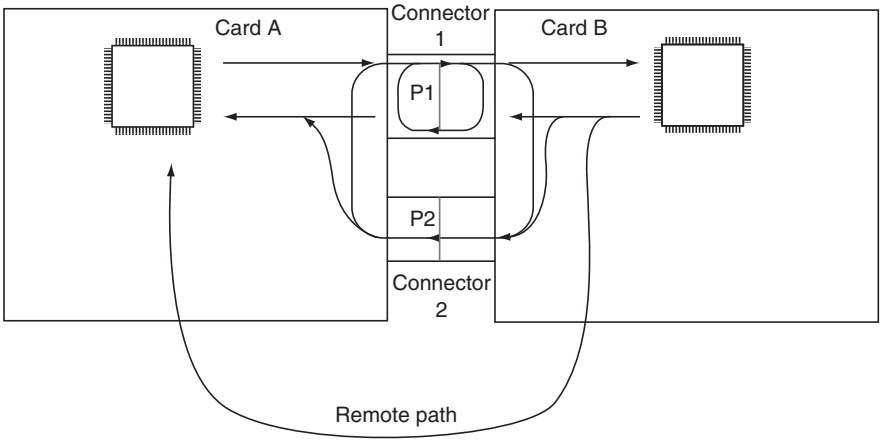


Figure 14.38 | EMI due to large current loops while using connectors.

EMI. High-frequency currents (especially from digital circuits) flowing in large loops will radiate lots of electromagnetic energy and will in general not pass the standard radiated emission tests. Therefore, it is highly important to contain all signal flow to loops having small cross-sections.

Referring to Figure 14.38, a portion of the signal currents flowing from card A to card B through connector 1 will return through the ground pins of the connector 1. This results in the loop path P1 as indicated in Figure 14.38. Likewise, a small fraction of the signal currents will return to card A through some ground pins of connector 2, which will result in the loop path P2. A still smaller fraction of the signal currents will return to card A, taking a rather circuitous loop as shown in the figure.

Consider that there are only two connectors as shown in Figure 14.38. The portion of the high-speed returning signal current that flows through connector 2 depends on the ratio of the inductance of loop path P1 to the inductance of the loop path P2. At very low frequencies, the amount of the current returning through the connector 2 depends on the ratio of the resistances. At higher frequencies, it is determined by the ratio of the inductances. As EMI is a high-frequency problem (> 30 MHz), one need only consider the inductance ratio of the two loops. Above 30 MHz, the standard limits on radiation are approximately 100 μV/m measured at a distance of 3 m from the equipment. A simple method to estimate the radiated electric field is as follows:

$$E = 1.4\times10^{-18}\left(\frac{AI_{\mathrm{p}}f_{\mathrm{c}}}{t_{\mathrm{r}}}\right) \tag{14.70}$$

where E the radiated electric field (in V/m) at 3 m distance; A the radiating loop area (inches^2); I_{p} the peak current in the loop (A); t_{r} the signal rise time (s); f_{c} the frequency of the loop current (Hz).

Some useful rules for reducing connector emissions are given as follows:

1. Use more grounds in connector 1. This will bring the grounds closer to each signal and lower the effective radiating loop area in connector 1, that is, reduce the loop path area A_{P1}.
2. Adding more grounds in connector 1 also lowers the inductance of the connector 1. This reduces the current flow in the remote loops.

3. Increase series inductance of the remote paths. This is done by using common mode chokes. The common mode chokes are connected to enclose the signal and return lines of connector 1. This will not provide any inductance for return paths through connector 1. However, these common mode chokes will act as differential chokes for all other remote return paths, thereby decreasing the currents through the remote paths.
4. All connectors on the mother boards should be as close as possible, so that the loop area of the remote return paths is reduced.
5. If possible, slow down the signal rise and fall times.

14.9 Part Derating

Derating can be defined as the operation of a part at less severe stresses than those for which it is rated. In practice derating can be achieved by

1. reducing stresses (e.g., snubber circuits as discussed in Chapter 2);
2. increasing the strength of the part.

Selecting a part of greater strength is usually the most practical approach. Derating is very effective because the failure rate of most parts tends to decrease as the applied stress levels are decreased below the rated value. Conversely, the failure rate increases when a part is subjected to higher stresses and temperature. Derating is one of the methods that is used in the iterative design process to improve the reliability. The initial design should incorporate default derating for the various components. Later the actual derating for the component should be adjusted based on the overall reliability requirement and the cost requirement. Component derating is definitely an important aspect of the design process, but it will improve only with experience. The reliability prediction process that will be discussed in the next chapter will provide the reliability values of the various components and modules. This too will act as an indicator to derate the components satisfactorily.

Electronic parts are prone to premature failure due to thermal overstress. Decreases in failure rate can be achieved in these cases by reduction of stress with adequate thermal design. Derating procedures vary with different types of parts and their application. Resistors are derated by decreasing the ratio of operating power to rated power; capacitors are derated by maintaining the applied voltage at a lower value than the voltage for which the part is rated. Semiconductors are derated by keeping the power dissipation below the rated level.

As a general rule, derating should not be conservative to the point where costs rises excessively. Neither should the derating criteria be so loose as to render reliable part application ineffective. Optimum derating occurs at or below the point on the stress/temperature curves where a rapid increase in failure rate is noted for a small increase in temperature or stress.

CONCLUDING REMARKS

The term "reliability" has always been used in a subjective sense to indicate the trustworthiness of a product or system. However, in order to design a system for a specified trustworthiness, one should be able to quantify reliability. As reliability encompasses both the functional and life aspects of the system. The functional aspects (intended function) is deterministic and the life aspect (perform the intended function for a specified period of time) is in a sense uncertain. Therefore, reliability is a probabilistic measure of the trustworthiness of the system.

In designing a product for a specified reliability, the life aspect is specified in terms of the failure rate or the MTTF of the product. The issue that arises after design is with regard to validation. Will the designed product meet the life requirement that is specified? How does one ascertain that the life specifications are incorporated? Evidently one cannot wait till the system fails. Validation is performed by reliability prediction. However, before even performing reliability prediction, one needs to model the system from the reliability point of view in order to apply the reliability prediction rules.

This chapter discusses essentially the methods involved in expressing a given system in terms of the reliability models such that they are amenable for mathematical analysis and reliability prediction. At the circuit level, all the devices of the circuit are in general essential and therefore, the series model is predominant at this level. At a still higher level, probably the board level, parallel model may prevail in some cases. Redundant models are prevalent only at the sub-system and system levels.

TUTORIAL EXERCISES

1. Consider a building that is powered from the mains grid. Some of the equipments are essential items that should not face a power shutdown. Therefore, these equipments are provided with alternative power sources. A battery-inverter system is one such alternative power source. A diesel generator set is another alternative power source. The essential equipments are normally powered by the mains grid. The diesel generator set is maintained in cold standby. The battery-inverter system is maintained in hot standby. The moment the grid power fails, the sense/switch network isolates the grid and connects the battery-inverter system that is in hot standby thereby continuing to supply power to the essential load. The battery-inverter system is rated to supply power for only 15 min. Before a maximum time of 15 min, the diesel generator system starts-up, becomes operative and initiates the sense/switch network to change over the load from the battery-inverter system to the diesel generator system. Using variables for the failures rates of the various sub-systems, obtain a relation for the reliability of the uninterruptible power system.

2. In some cases, especially for new components, neither the base failure rate model nor the nomo-graphs are available to estimate the failure rate of the component. In such cases, the failure rate has to be estimated by conducting experiments on a population of the device by inducing failures through accelerated tests. After the failure data for the population is tabulated, the device may be characterized either by non-parametric analysis or parametric analysis. In non-parametric analysis, the statistical parameters like mean and standard deviations are used to characterize the device failure. In the parametric analysis, a curve fit to exponential distribution is performed to extract the parameters like failure rate. In this regard, conduct a study on the following:
   - **a.** How are accelerated tests conducted?
   - **b.** Where is Arrhenius model used in accelerated test experiments?
   - **c.** Study the methods of performing non-parametric analysis.
   - **d.** Study the methods of performing parametric analysis.
   - **e.** What is goodness of the curve fit?

FILL IN THE BLANKS

1. A product may be said to be a compromise among the three primary constraints _____, _____ and _____.
2. The various electronic components are subjected to two types of stresses that are _____ and _____.
3. Reliability is the _____ that an item will perform its intended function for a specified period of _____ under the stated conditions.
4. T is a continuous random variable that denotes the time at which _____ of an item occurs.
5. The reliability of an item can be mathematically defined as _____.
6. Reliability has been described as quality in the _____ dimension.
7. Exponential distribution is a special case of Weibull distribution when the Weibull parameter, β is _____.
8. Infant mortality period is also called _____ period.
9. The failure rate _____with time during the infant mortality period.
10. The failure rate is _____ with respect to time during the useful life period.
11. The failure rate _____ with time during the wear out period.
12. Useful life period is characterized by the occurrence of _____ failures.
13. The failure-time probability density function of the system during the useful life period can be modeled as an _____ distribution.
14. In a series reliability model, _____ components must work to ensure system success.
15. In a parallel reliability model, _____ component must work to ensure system success.
16. The switching network _____ and _____ the failed component.
17. In dormant-mode failure of the sense/switch network, the sensing mechanism fails and the system _____ to work till the _____ of one of the components of the parallel redundant system.
18. In immediate failure mode of the sense/switch network, failure in the switching network leads to _____ failure of the system.
19. In standby redundancy with multiple failure the occurrence of failure among the m operational units is a _____ process.
20. The main objective of the design process is to combine the various _____ and constraints in an appropriate manner to result in the _____ of the product.
21. Larger number of connectors implies _____ reliability.
22. Connectors present a classic case of trade-off between _____ and _____.
23. Failure rate of most parts tends to _____ as the applied stress levels are decreased below the _____ value.
24. The failure rate of a part _____ when a part is subjected to higher stresses and temperature.

DESCRIPTIVE QUESTIONS

1. Discuss the general design process of a product.
2. Discuss the relationships between $R(t)$, $F(t)$ and $Z(t)$.
3. What is failure rate?
4. Failure rate or the hazard rate completely determines the failure-time probability density function. Discuss.
5. What is the Weibull model of the failure rate? Explain under what conditions the failure rate becomes constant or increasing with respect to time or decreasing with respect to time?
6. Plot and discuss the failure-rate curve during the life of an electronic equipment.
7. Write short notes on: (a) infant mortality, (b) useful life and (c) wear-out period.
8. Discuss the relationship between failure rate, mean time to failure and mean time between failures.
9. Write short notes on: (a) series system, (b) parallel system and (c) meshed system.
10. Discuss the methods for evaluating the reliability of mesh systems.
11. What is conditional probability approach towards evaluating the reliability of mesh systems?
12. What is the limitation of the multiple-path equivalent method for evaluating mesh systems?
13. What is the motivation for using redundant systems?
14. What are the different redundant topologies?
15. What is the function of switching networks in redundant systems?
16. Write short notes on: (a) dormant mode failure and (b) immediate mode failure.
17. What is hot standby?
18. What is the effect of non-ideal sense/switch on the MTBF of the system?
19. Discuss the method used to evaluate the reliability of non-identical components.
20. What are the factors affecting the reliability of a system?
21. What are target specifications? What are its uses?
22. Explain the stimuli and responses for a system.
23. Discuss segregation of modules into sub-modules.
24. Discuss the inter-connection and the naming convention for PCB and non-PCB interconnections.
25. What is reliability apportionment and what is its importance?
26. Discuss the various methods of reliability apportionment.
27. List the connection types in the ascending order of failure rates.
28. What are the methods of reducing the crosstalk between high-speed signal pins?
29. Explain the problems imposed by connectors in high-speed digital circuits.
30. What is parts derating? Why and where is it used?
31. What are the methods of reducing the stresses within a device?

PROBLEMS

1. A system design requires 100 identical components in series. If the overall reliability must not be less than 0.98, what is the minimum reliability of each component?
2. A system's reliability is 0.9 at the end of 5 years. What is the mean time to failure of the system? If the system comprises 10 identical and essential sub-components, then what is the failure rate of each sub-component?
3. A system design requires 100 identical components in series. If the overall reliability must not be less than 0.98, what is the minimum reliability of each component?
4. A system's reliability is 0.95 at the end of 2 years. What is the mean time to failure of the system? If the system comprises of 10 identical sub-components, then what is the failure rate of each sub-component?
5. Consider the reliability block diagram shown in Figure 14.39. Evaluate the reliability of the system.
6. Consider the system shown in Figure 14.40 wherein success requires that at least one path exists from point X to point Y. Here the component E is a unidirectional component. If each component has a failure probability of q, evaluate the reliability of the system.

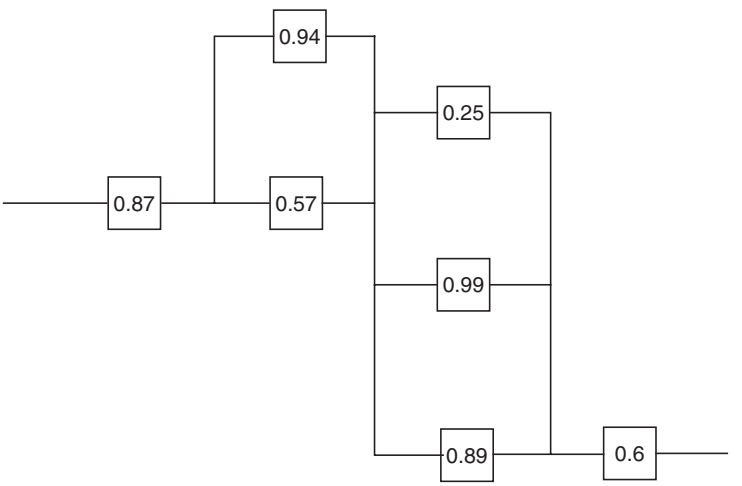


Figure 14.39 | Problem 5.

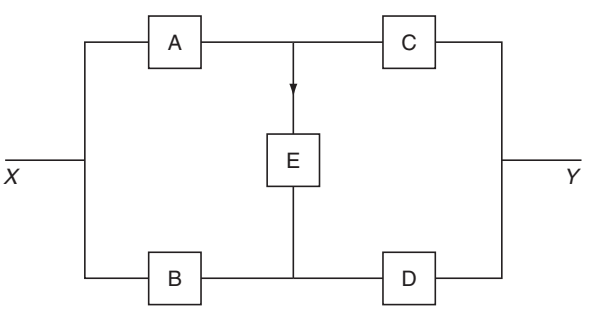


Figure 14.40 | Problem 6.

7. Derive the relationship between MTBF and reliability.
8. Derive the relationship between MTBF and failure rate for a *m*-survivor full-on redundant configuration.
9. A system has three units having identical failures rates that are connected as 2-survivor full-on redundant configuration. If the reliability of the overall system is 0.985 at the end of 2 years, then what should be the minimum reliability of each redundant unit? What is the failure rate of each unit?
10. Derive the reliability equation for the *m*-survivor full-on redundant configuration with imperfect sense/switch network.
11. A system comprises two identical units which are connected in full-on, single-survivor redundant configuration. If the failure rate of one of the units is 0.001 f/h, what should be the minimum reliability requirement for the switching network, with regard to dormant failures, for a period of 5 h of continuous operation, so that the redundant system has a reliability greater than the individual units? (Assume that the switching network does not give rise to immediate failure of the system).
12. For Problem 11, if the probability of immediate mode failure is 0.05, then what should be the probability that the sense/switch mechanism does not fail in dormant mode such that the overall reliability of the system is greater than the individual units.
13. Derive the relationship between MTBF and failure rate for an *m*-survivor standby redundant configuration.
14. Derive the reliability equation for the *m*-survivor standby redundant configuration with imperfect sense/switch network.
15. Consider a system wherein there are five sub-systems. The guess estimate of the failure rates of these five sub-systems are 10 FITs, 100 FITs, 140 FITs, 45 FITs and 200 FITs respectively. If the reliability requirement from the target specification is 700 FITs, then using the ARINC method of reliability apportionment, obtain the failure rate requirements for the five sub-systems.
16. Two I/O pins of a microcontroller are buffered and connected to another printed circuit board by means of a connector. The two signals are connected to adjacent pins of the connector. On either side of the two signal pins are two ground pins. The distance between any two pins of the connector is 2 mm and the diameter of the pin is 0.8 mm. The rise time of the digital switching currents flowing through the buffered signal pins is 0.1 A/μs. Estimate the mutual inductance between the two signal pins. Estimate also the crosstalk.
17. For the Problem 16, if the ground pins are in between the two signal pins, then estimate the crosstalk.

ANSWERS

Fill in the Blanks

1. performance; cost; life
2. electrical; temperature
3. probability; time
4. failure
5. $R(t) = 1 - F(t)$
6. time
7. unity
8. burn-in
9. decreases
10. constant
11. increases
12. stress-related

13. exponential
14. all
15. any
16. detects; isolates
17. continues; failure
18. instantaneous
19. Poisson
20. demands; conception
21. lesser
22. reliability; maintainability
23. decrease; rated
24. increases

CHAPTER 15

Reliability Prediction

Learning Objectives

After reading this chapter, you will be able to:

- understand and apply part stress co-variate method for reliability prediction.
- iterate the design to achieve the specified reliability goals.

Reliability prediction is the process of quantitatively assessing whether a proposed or actual equipment/system design will meet a specified reliability requirement. Reliability prediction does not contribute to system reliability but does constitute decision criteria for selecting courses of action which affect reliability.

The primary objective of reliability prediction is to provide guidance, relative to expected inherent reliability of a given design. Reliability predictions are most useful and economical during the early phase of a system design, before the hardware is constructed and tested. During design and development, predictions serve as quantitative guides by which design alternatives can be judged for reliability. Basically, the purpose of reliability prediction includes feasibility evaluation, comparison of alternative configurations, identification of potential problems during design review, logistics support planning and cost studies, determination of data deficiencies, trade-off decisions and apportioning/allocation of requirements.

The prediction of the reliability of the electronic product has the following uses:

1. Determination of reliability requirements in planning, preliminary design specifications and determination of the feasibility of proposed reliability requirement. Comparison of the determined reliability requirement with state of the art, feasibility for guidance in budget and schedule decisions.
2. Identification and ranking of potential problem areas and the suggestion of possible solutions.
3. Allocation of reliability requirements among the sub-systems and lower level items.
4. Evaluation of the choice of proposed parts, materials, units and processes.
5. Evaluation of the design for prototype fabrication during the development phase.
6. Provides a basis for trade-offs and compromises.

Therefore, reliability prediction is a key to system development and allows reliability to become an integral part of the design process. To be effective, the prediction technique must relate engineering variables to reliability variables. In general, reliability prediction depends on the

1. depth of knowledge of the design;
2. availability of historical data

In the initial stages of development, the data available is qualitative which progressively becomes quantitative at later stages. Therefore, there are reliability prediction techniques that can be used at various stages of the product development.

15.1 Similar Equipment Technique

The equipment under consideration is compared with similar equipments of known reliability in estimating the probable level of achievable reliability. This is one of the simplest technique used in the early stages of the system design which involves a simple estimate of the equipment reliability in terms of mean time between failures (MTBF), failure rate, etc. based on experience gained from similar operational equipments with similar function.

EXAMPLE 15.1 If the reliability of a previously designed DSP board is available, then another DSP board with DSP from another manufacturer and similar features would result in a new product with reliability in the same order of magnitude as the old board.

15.2 Similar Complexity Technique

The reliability of a new design is estimated as a function of the relative complexity of the subject item with respect to a typical item of similar type. Generally, a nomo-graph between MTBF and the number of active components in an electronic product is used for the initial reliability prediction. The nomo-graph between MTBF and the number of active components is developed from past design experience with various products. A plot of the MTBF of the various products with respect to the number of active components can be used as a graphical tool for predicting reliability. In order to have a better estimate of the reliability, a family of plots of the MTBF versus number of active components with active device rating as a parameter could be developed.

EXAMPLE 15.2 Consider that past history exists in terms of specification, achieved reliability and number of active devices for a 5 KVA UPS system. If the requirement is to design a 50 KVA UPS system, then it may be deduced that the design complexity is similar to the 5 KVA UPS system; however, the active devices will handle different electrical and thermal stresses. The nomo-graph of the MTBF versus active device rating may be a helpful guide to estimate the pre-design reliability of the system.

15.3 Prediction by Function Technique

This refers to a prediction technique which relates the expected reliability with the functional characteristics of the equipment or sub-system. This is also an approximate method of reliability prediction that is useful in pre-design stages of the equipment development. The technique is based upon a statistical correlation between significant functional characteristics and the observed operational reliability of an equipment.

The result is a series of regression equations which relate the more significant equipment functions (from the reliability viewpoint) to the expected reliability.

EXAMPLE 15.3

Consider the case of an analog-to-digital converter (ADC) card. The mean time to failure (MTTF) can be related to the most significant functional specifications as follows:

$$\text{MTTF} = f(\text{number of channels, resolution, accuracy, sampling time}).$$

This relationship is again obtainable from past history or experience. The MTTF as a function of various significant functional specifications is similar to a curve-fit algorithm to statistical data.

15.4 Parts Count Technique

This technique is used when one has a feel for the number of component parts by class or type that will be used in the equipment but does not have enough data with respect to the stress that each part will be subjected in the final design. It involves counting the number of parts of each class or type, multiplying this number by the generic failure rate for each part class and summing these products to obtain the failure rate for the equipment.

This technique of parts count analysis is generally used in the preliminary design stages where the general circuit diagram is fixed but the values of the circuit components are not yet determined. It gives the worst case value for failure rate for a given number of components assuming series model, that is, all the components must work for satisfactory performance of the circuit.

The general expression for the item failure rate is

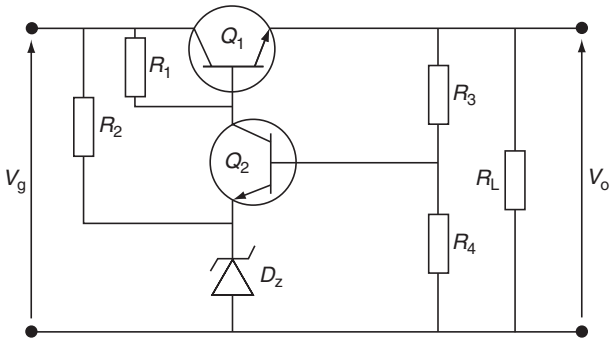
$$\lambda_{\text{item}} = \sum_{i=1}^{n} N_i (\lambda_G \pi_E \pi_Q)_i$$

where λ_{item} is the total item or equipment failure rate; λ_{Gi} the generic failure rate of the ith generic part; π_E the environment adjustment factor or environment co-variate; π_Q the quality adjustment factor or quality co-variate; N_i the quantity of the ith generic part; n the number of different generic part categories. *The failure rates λ_{item} and λ_G are expressed in terms of failure units called FITs. FITs is a sort of acronym for Failure unITs. 1 FIT means 1 failure in 10^9 hours.*

There are many standards in use to determine the co-variates and the generic or the base failure rates of the items. Typical standards are the MIL-217 of US military, the BELL (BELCORE) standards for commercial and communication equipments from AT & T Bell, HRD standards from British Telecom and the French CNET standards. The MIL-217 is one of the most popular standards used worldwide. This chapter illustrates the reliability prediction examples using the MIL-217 standards. The relevant failure rate data are available in the MIL-HAND BOOK. It should, however, be noted that whichever standard is being used, the same standard should be used as a platform for comparing various other products from the reliability point of view.

Problem 15.1

For the series-pass linear regulator shown in Figure 15.1, perform a parts count reliability prediction.



V_g = 100 V, 10 A
V_{out} = 75 V, 10 A
Q_1 dissipates around (100 − 75) × 10
= 250 W

Figure 15.1 | Problem 15.1.

Solution

Table 15.1 | Parts count failure rate estimation

Components	N	$\lambda_{item} = \sum_{i=1}^{n} N_i(\lambda_G \pi_E \pi_Q)$ FITs	$N\lambda$ FITs	$R(t)$
Resistors: 1/4 W cc	5	27.1	135.5	$e^{-135.5t}$
Transistors: High power	1	16.68	16.68	$e^{-16.68t}$
Low power	1	3.92	3.92	$e^{-3.92t}$
Diode: Zener	1	21.68	21.68	$e^{-21.68t}$
			177.78	

$$\lambda_{total} = \Sigma N\lambda = 177.78 \text{ failures}/10^9 \text{ hours}$$

$$R(t) = (e^{-135.5t})\ (e^{-16.68t})\ (e^{-3.92t})\ (e^{-21.68t}) = e^{-177.78t}$$

where a unit time t represents 10^9 h.

Note: The above data can be obtained from the MIL-HANDBOOK or by executing the MATLAB-based toolbox for reliability prediction [Reliability for Electronic Circuit (REC)] given in the accompanying CD. The failure rates are calculated for 50% temperature and electrical stresses. The environment considered is ground benign (GB) and the quality considered is of quality type "LOWER".

Part Count with Non-Series Model

In case if some of the components are non-series in the reliability block diagram, those can be calculated using techniques discussed in evaluating non-series blocks as discussed in the Section 14.2. They can be treated as a module (unit). One cannot simply add the failure rates as the equivalent non-series block system will not have an exponential distribution. In such cases, simply multiply the reliabilities of all generic parts including the non-series modeled unit.

$$R(t) = R_1(t)R_2(t)\ldots R_n(t)$$

The MTBF can be evaluated using

$$\text{MTBF} = \int_0^\infty R(t)\mathrm{d}t$$

15.5 Part Stress Technique

The equipment failure rate is determined as an additional function of all part failure rate and operational stress levels and derating characteristics of each part. Here the stresses in the components are also taken into account. This is the most accurate method of reliability prediction that is currently in use.

The part stress technique is very much similar to the part count technique except that the thermal and electrical stress are evaluated for each part and accounted for in the failure rate model as temperature and electrical stress co-variates. Here

1. Base failure rate for each part has to be determined. This is determined by failure rate models, appropriate tables or graphs.
2. The values of the adjustment factors or co-variates are to be determined.
3. The part failure rate is calculated using the base failure rate and the co-variates.

The remaining procedure is similar to the parts count technique where the series model is assumed. Here the item failure rate is given by

$$\lambda_{\text{item}} = \sum_{i=1}^{n} N_i(\lambda_G \pi_E \pi_Q \pi_S \pi_T)_i$$

where there are two additional co-variate factors: π_S the electrical stress factor and π_T the temperature stress factor as compared to the parts count method.

Problem 15.2

Consider the three-phase full-bridge circuit as shown in Figure 15.2. The inverter is used to drive a three-phase motor which has the following ratings.

Motor rating = 3 kW
Power factor, pf = 0.8
Efficiency, $\eta = 0.9$

Evaluate the reliability of the MOSFET bridge. The MOSFETs are known to handle 20 W of power dissipation and have an ON-resistance of 0.1 Ω. The rise and the fall times can together be taken to be 1 μs. The V_{dclink} voltage is 600 V and the switching frequency of the inverter is 20 kHz. However, it should be

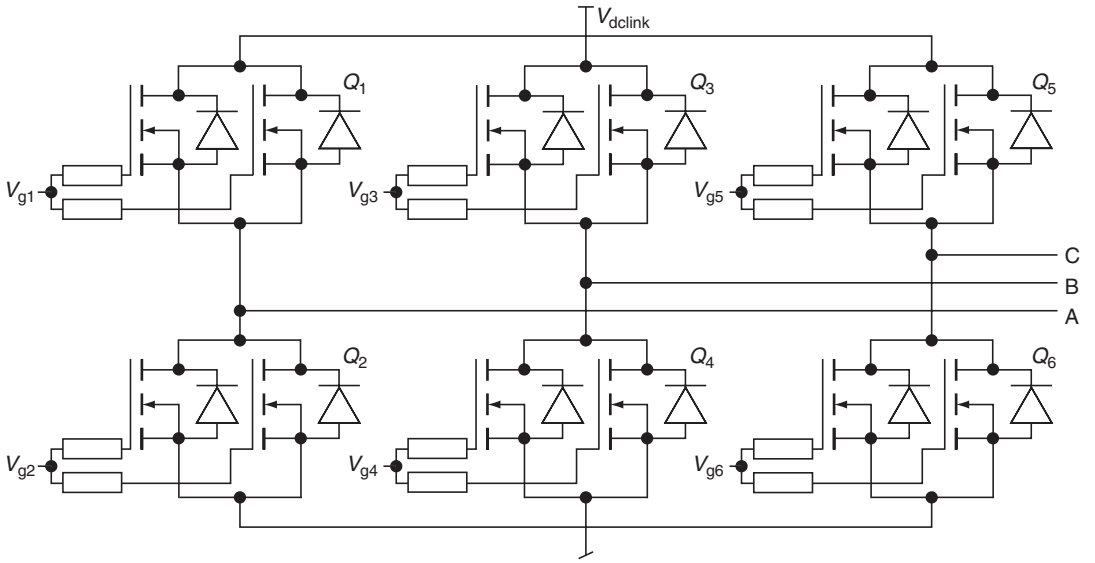


Figure 15.2 | Problem 15.2.

noted that if space-vector modulation is used, the frequency at which the MOSFETs of the inverter will switch is half of 20 kHz, that is, 10 kHz.

Note: The inverter bridge circuit shown in Figure 15.2 is implemented in practice with IGBTs especially for grid voltages. This example is only representative and used to illustrate the reliability concepts.

Solution

$$\text{VA} = \frac{3\text{ kW}}{0.8\times 0.9} = 4.166\text{ kW}$$

$$\text{VA} = \sqrt{3}V_{\text{L}}I_{\text{L}}$$

$$I_{\text{L}} = \frac{VA}{\sqrt{3}\times 400} = 6\text{ A rms}$$

The peak device current = $\sqrt{2}\times 6 = 8.5\text{ A} = I_{\text{m}}$

The current through each switch is a half sinusoid and therefore the root mean square (rms) current through each switch is approximately $I_{\text{m}}/2$ (neglecting the small periods during diode freewheeling). If each inverter switch consists of single MOSFET device, then the conduction loss is given as

$$P_{\text{cond}} = I_{\text{ms}}^2 R_{\text{Dson}} = \left(\frac{I_{\text{m}}}{2}\right)^2 R_{\text{Dson}} = \left(\frac{8.5}{2}\right)^2 \times 0.1 = 1.8\text{ W}$$

Each MOSFET is switching at a frequency of 10 kHz.

$$P_{\text{switching}} = \left(\frac{V_{\text{dclink}} I_{\text{aV}} f(t_{\text{r}} + t_{\text{f}})}{2} \right) \times 0.5$$

(0.5 is used here because only half sine wave flows through each MOSFET in a fundamental period).

$$I_{\text{aV}} = \frac{2I_{\text{m}}}{\pi} = 5.4 \text{ A}$$

Using the appropriate values, $P_{\text{switching}} = 16.2$ W.

$$P_{\text{loss}} = 1.8 + 16.2 = 18 \text{ W per MOSFET}$$

$$P_{\text{loss}} \text{ rated} = 20 \text{ W}$$

From the reliability tables of MIL-Handbook or from the accompanying REC toolbox, the failure rate for the MOSFET can be computed. The environment is considered as *ground benign*, and the quality level is considered as *lower*. The heat sinks for the MOSFETs are selected such that the operating junction temperature is 100°C for a maximum ambient of 50°C.

The failure rate for each MOSFET is calculated as

$$\lambda_{\text{item}} = 1100 \text{ FITs}$$

If for each inverter switch, two MOSFETs are connected in parallel as shown in Figure 15.2, then:

$$P_{\text{cond}} = \left(\frac{8.5/2}{2} \right)^2 \times 0.1 = 0.45 \text{ W}$$

$$P_{\text{switch}} = 8.1 \text{ W}$$

$$P_{\text{loss}} = 8.55 \text{ W per MOSFET}$$

The calculated failure rate is now

$$\lambda_{\text{item}} = 743.94 \text{ FITs}$$

The reliability of the parallel combination of two MOSFETs is given by (see Chapter 14)

$$R_{\text{sw}} = e^{-2\lambda_2 t} + \left(\frac{2\lambda_2}{\lambda_1 - \lambda_2 - \lambda_{02}} \right) (e^{-(\lambda_2 + \lambda_{02})t} - e^{-\lambda_1 t})$$

where λ_1 is the failure rate of the MOSFET when it is sharing the 100% load; λ_2 the failure rate of the MOSFET when it is sharing 50% load during parallel operation; λ_{02} the probability that the MOSFET fails in open-circuit mode during parallel operation = 10% of λ_2. Statistically it is found that 10% of the time, the MOSFET initially fails in open-circuit failure mode.

Thus,

$$R_{\text{sw}} = e^{-1488t} + 5.2841(e^{-818.4t} - e^{-1100t})$$

There are six such paralleled MOSFET units in the bridge which are considered as series units from the reliability point of view. Therefore, the overall reliability of the MOSFET bridge would be

$$R_{\text{overall}} = R_{\text{sw}}^6 = [e^{-1488t} + 5.2841(e^{-818.4t} - e^{-1100t})]^6$$

The MTTF is obtained by integrating the reliability R_{overall} from 0 to ∞.

15.6 Environments

While predicting the reliability of the products, the environment in which the product ultimately operates has a significant bearing on the overall MTTF of the product. All reliability standards define many standard environments in which a particular product is designed to operate. Each environment type is associated with a symbolic notation indicating the type of standard environment. The MIL-STD defines the following types of standard environments and their associated notations:

1. Ground benign, G_B.
2. Ground fixed, G_F.
3. Ground mobile, G_M.
4. Naval sheltered, N_S.
5. Naval unsheltered, N_U.
6. Air-borne cargo, A_{IC}.
7. Air-borne fighter, A_{IF}.
8. Air-borne unmanned cargo, A_{UC}.
9. Air-borne unmanned fighter, A_{UF}.
10. Air-borne winged, A_{RW}.
11. Space flight, S_F.
12. Missile flight, M_F.
13. Missile launch, M_L.
14. Cannon launch, C_L.

15.7 Reliability Prediction Process

Reliability prediction involves the use of nomo-graphs, tables, curve-fit data models and failure rate equations. Therefore, it is cumbersome to calculate the reliability even for a simple circuit, let alone a complex one, by manual process. There is a need for reliability evaluation and design tool to aid the iterative system design implementation. This section formalizes the framework of the reliability prediction process that can be implemented on program environments like C, MATLAB or SciLAB. However, the reliability prediction toolbox included in the accompanying CD is with respect to the MATLAB environment.

The design of a circuit with reliability aspects integrated into the design is a three-step process. The following steps illustrate the circuit-design process:

Step 1: Build the Circuit Template

After finalizing the circuit to be designed, generate a circuit template file that contains all the components that go into the circuit. Every component is treated as an object comprising electrical and reliability based variables and constants. The component naming convection and the object definition is followed as per the discussion in the next section.

Step 2: Perform Functional Design using the Circuit Template

Based on the target specifications of the circuit, the functional design of the circuit is performed. From the functional design parameters, the reliability aspects like the temperature stress due to the power dissipation in the component and the electrical stresses should be computed to complete the definition and selection of a particular component object.

Step 3: Display Results

The design of the circuit should be evaluated by a computer program and the target specifications, functional design parameters, component selection parameters and the failure rate estimates are displayed.

If the predicted MTTF or MTBF value for the circuit does not satisfy the circuit's life specification, then one or more of the following actions may be performed in sequence,

1. Without changing the topology of the circuit, derate the component or components that contribute the most to the overall failure rate of the circuit.
2. Modify portions of the circuit that contribute to large failure rate by simplifying the circuit or using de-stressing circuits like paralleling of MOSFETs and snubber circuits.
3. If it is not possible to meet the reliability specifications by 1 and 2, then use the redundant topologies as discussed in the previous chapter.

It should be noted that at the circuit level, *all the components are essential, which implies that circuits are series models from the reliability point of view*. The exception to this are circuits like paralleling of MOSFETs and paralleling of diodes as discussed in examples in the previous chapter. The redundant configurations are used at a higher level of system integration like sub-system level or even at the system level.

15.8 Component Naming Convention

The following is the list of generic electronic components used in the reliability analysis and reliability prediction toolbox (REC). The first letter for a particular component variable is pre-defined (also in the toolbox) and is indicated in the list given as follows:

1. BJT, Q.
2. Capacitor, C.
3. Connections, SC.
4. Connectors, J.
5. Diode, D.
6. Electronic filters, EF.
7. Fuses, FU.
8. Inductance, L.
9. Lamps, LA.
10. Laser, LAS.
11. Meters panel, MEP.
12. Microcircuits, U.
13. Opto-electronics, OPT.
14. Potentiometer, POT.
15. Quartzcrystals, QZ.
16. Resistor, R.
17. Relays, RY.
18. Rotating device, ROT.

19. Switches, SW.
20. Transformer, XFM.
21. Transistor-FET, M.
22. Tubes, TUB.
23. Thyristors, SCR.
24. Uni-junction transistor, UJT.
25. Variable capacitor, VC.
26. Miscellaneous, MIS.
27. HYBRID micro, HYB.
28. User defined, USR.

The failure rate or hazard rate is represented as "Hp". In the case of hybrid packages, the components within the hybrid package are made up of the standard discrete components. To indicate that the components like say a BJT is used in the hybrid package, the notation is the component notation followed by 'H', that is, QH. For capacitance used in the hybrid package, the notation is 'CH'. Thus an 'H' is appended to the normal component notations if they are used within a hybrid package.

15.9 Component Object Notation

An electronic component as mentioned earlier is an object comprising both electrical and reliability parameters. Specifying all these parameters would define the specific component. Each category of component, say a resistor, may contain many types of resistors like film, carbon compound, etc. Each type may require different parameters to define the resistor object. In the object notations, among the parameters, the more common parameters are: "type" used to indicate the device type, "PE_type" used to indicate the environment type, "PQ_type" used to indicate the quality type, "PA_type" used to indicate the application type, "Tj" used to indicate the operating junction temperature, "Vr" used to indicate the rated voltage, "S" used to indicate the electric stress ratio and "Pr" used to indicate the rated power. "Hp" is used to indicate the part hazard rate or the failure rate. All the object parameters for the various components are described in the Appendix VI. The component object notation is given as follows, using an illustrative example:

$$Q(1).\text{type}$$

where Q is the device symbol; 1 the device index and type the parameter.

The first letter is the device symbol as discussed in Section 15.8. The number within the parenthesis is the device index. It can be any integer that uniquely identifies the device within the device class. The parameter that needs to be quantified is indicated following the period (.). The REC toolbox also follows the component object notation discussed in this section. It generates the circuit-design template file based on this component object notation.

It should be noted that all the parameters within the component object should be defined either by a relevant constant or an equation based on the circuit specification in order to wholly specify the component. Based on such a component specification, a component with appropriate rating from any particular manufacturer may be selected. A sample list of some major component objects is given below.

(Refer to Appendix VI for a complete description of the component parameters.)

```
BJT component object
Q(1).type =
Q(1).PE_type =
Q(1).PQ_type =
Q(1).Tj =
Q(1).VCE =
Q(1).Vr =
Q(1).Pr =
Q(1).PA_type =
Q(1).Hp=

Capacitor component object
C(1).type =
C(1).PE_type =
C(1).PQ_type =
C(1).T =
C(1).S =
C(1).C =
C(1).Hp=

Connection object
SC(1).type =
SC(1).PE_type =
SC(1).PQ_type =
SC(1).N1 =
SC(1).N2 =
SC(1).P =
SC(1).Hp=

Connector object
J(1).type =
J(1).PE_type =
J(1).PQ_type =
J(1).C =
J(1).TA =
J(1).G =
J(1).i =
J(1).Hp=

Diode component object
D(1).type =
D(1).D_type =
D(1).PE_type =
D(1).PQ_type =
D(1).Tj =
D(1).PC =
D(1).Vo =
D(1).Vr =
D(1).Hp=

Electronic flter object
EF(1).PE_type =
EF(1).PQ_type =
EF(1).Hb_type =
EF(1).Hp=

Fuse object
FU(1).PE_type =
FU(1).Hp=

Inductor object
L(1).type =
L(1).PE_type =
L(1).PQ_type =
L(1).THS =
L(1).Hp=

Lamp object
LA(1).PE_type =
LA(1).PA_type =
LA(1).PU =
LA(1).Vr =
LA(1).Hp=

Laser component object
LAS(1).type =
LAS(1).PE_type =
LAS(1).HM_type =
LAS(1).HC_type =
LAS(1).Hp=

Meter panel object
MEP(1).PE_type =
MEP(1).PQ_type =
MEP(1).PF_type =
MEP(1).PA_type =
MEP(1).Hp=

Microprocessor object
U(1).PE_type =
U(1).PQ_type =
```

```
U(1).pkg =
U(1).y =
U(1).Np =
U(1).family =
U(1).N =
U(1).Tj =
U(1).Hp=

Linear IC object
U(2).PE_type =
U(2).PQ_type =
U(2).pkg =
U(2).y =
U(2).Np =
U(2).N =
U(2).Tj =
U(2).Hp=

Logic array object
U(3).PE_type =
U(3).PQ_type =
U(3).pkg =
U(3).y =
U(3).Np =
U(3).family =
U(3).N =
U(3).Tj =
U(3).Hp=

Memory component object
U(4).PE_type =
U(4).PQ_type =
U(4).pkg =
U(4).y =
U(4).Np =
U(4).family =
U(4).B =
U(4).MT =
U(4).C =
U(4).ECC =
U(4).Tj =
U(4).Hp=

VLSI component object
U(5).PE_type =
U(5).PQ_type =
U(5).Np =
U(5).pkg =
U(5).CFT =
U(5).Tj =
U(5).pt =
U(5).mpt =
U(5).VTH =
U(5).XS =
U(5).DA1 =
U(5).Hp=

Digital IC object
U(6).PE_type =
U(6).PQ_type =
U(6).y =
U(6).pkg =
U(6).Np =
U(6).family =
U(6).N =
U(6).Tj =
U(6).Hp=

Opto electronic component object
OPT(1).type =
OPT(1).PE_type =
OPT(1).PQ_type =
OPT(1).Tj =
OPT(1).Hb_type =
OPT(1).Hp=

Potentiometer object
POT(1).type =
POT(1).PE_type =
POT(1).PQ_type =
POT(1).R =
POT(1).S =
POT(1).T =
POT(1).Ntaps =
POT(1).R1 =
POT(1).Pa =
POT(1).VR =
POT(1).PC_type =
POT(1).Hp=

Quartz crystal component object
QZ(1).PE_type =
QZ(1).PQ_type =
QZ(1).F =
QZ(1).Hp=
```

```
Resistor component object
R(1).type =
R(1).PE_type =
R(1).PQ_type =
R(1).T =
R(1).S =
R(1).Pd =
R(1).Hp=

Relay component object
RY(1).type =
RY(1).PE_type =
RY(1).PQ_type =
RY(1).Hb_type =
RY(1).Hp=

Rotating device object
ROT(1).type =
ROT(1).T_type =
ROT(1).M_type =
ROT(1).T =
ROT(1).LC =
ROT(1).Hp=

Switch component object
SW(1).type =
SW(1).PE_type =
SW(1).PQ_type =
SW(1).PC_type =
SW(1).PL_type =
SW(1).S =
SW(1).Hp=

Transformer object
XFM(1).type =
XFM(1).PE_type =
XFM(1).PQ_type =
XFM(1).THS =
XFM(1).Hp=

MOSFET component object
M(1).type =
M(1).PE_type =
M(1).PQ_type =
M(1).Tj =
M(1).Hb_type =
M(1).Hp=

Tube component object
TUB(1).type =
TUB(1).PE_type =
TUB(1).T =
TUB(1).Hb_type =
TUB(1).PK_type =
TUB(1).Hp=

Thyristor component object
SCR(1).PE_type =
SCR(1).PQ_type =
SCR(1).Tj =
SCR(1).VBo =
SCR(1).VBr =
SCR(1).Ifrms =
SCR(1).Hp=

Unijunction transistor object
UJT(1).PE_type =
UJT(1).PQ_type =
UJT(1).Tj =
UJT(1).Hp=

Variable capacitor object
VC(1).type =
VC(1).PE_type =
VC(1).PQ_type =
VC(1).T =
VC(1).S =
VC(1).Hb_type =
VC(1).Hp=

Example Hybrid component object
%-------------HYBRID CIRCUIT PACKAGE BEGINS-------------

HYB(1).m_type = [1 4 5 6 7 8 10];
HYB(1).DA_type = 2;
HYB(1).DA =
HYB(1).PD =
HYB(1).TC =
HYB(1).cdn = 0;
```

```
HYB(1).Tj=Tj_hybrid(HYB(1).m_type,HYB(1).DA_type,HYB(1).DA,HYB(1).
  PD,HYB(1).TC,HYB(1).cdn);

QH(1,1).type = 1;
QH(1,1).Tj = HYB(1).Tj;
QH(1,1).VCE =
QH(1,1).Vr =
QH(1,1).Pr =
QH(1,1).PA_type = 2;
QH(1,1).Hp=hybrid_BJT(QH(1,1).type,QH(1,1).Tj,QH(1,1).VCE,QH(1,1).
  Vr,QH(1,1).Pr,QH(1,1).PA_type);

CH(1,2).type = 1;
CH(1,2).T = HYB(1).Tj;
CH(1,2).S =
CH(1,2).Hb_type = ;
CH(1,2).C =
CH(1,2).Hp=hybrid_capacitor(CH(1,2).type,CH(1,2).T,CH(1,2).
  S,CH(1,2).Hb_type,CH(1,2).C);

DH(1,3).type = 1;
DH(1,3).D_type = 1;
DH(1,3).Tj = HYB(1).Tj;
DH(1,3).PC = 1;
DH(1,3).Vo =
DH(1,3).Vr =
DH(1,3).Hp=hybrid_diode(DH(1,3).type,DH(1,3).D_type,DH(1,3).
  Tj,DH(1,3).PC,DH(1,3).Vo,DH(1,3).Vr);

MH(1,4).type = 2;
MH(1,4).Tj = HYB(1).Tj;
MH(1,4).PA_type = 1;
MH(1,4).P =
MH(1,4).F =
MH(1,4).PM_type = 3;
MH(1,4).Hp=hybrid_FET(MH(1,4).type,MH(1,4).Tj,MH(1,4).PA_
  type,MH(1,4).P,MH(1,4).F,MH(1,4).PM_type);

HYB(1).CKT = 5;
HYB(1).PE_type = 1;
HYB(1).class = 16;
HYB(1).Y =
HYB(1).Hp_=hybridmicro(HYB(1).CKT,HYB(1).PE_type,HYB(1).
  class,HYB(1).Y);
Hc = [QH(1,1).Hp CH(1,2).Hp DH(1,3).Hp MH(1,4).Hp ];
HYB(1).Hp = sum(Hc)*HYB(1).Hp_;

%------------HYBRID CIRCUIT PACKAGE ENDS-------------
```

15.10 Example Reliability Prediction Process

Steps to Design a Circuit

1. Generate the *design template* for the circuits.
2. Enter the specifications and the design equations for all the parameters of all the component objects.
3. Execute the program to compute the functional parameters and the reliability parameters and display the results.
4. If the MTTF of the system is not satisfactory, then iterate the design by appropriately changing the rating.

Problem 15.3

A Simple BJT Base Drive: *Design/select the components of the circuit in Figure 15.3 according to the functional specification given below and predict the reliability after 1 year of operation.*

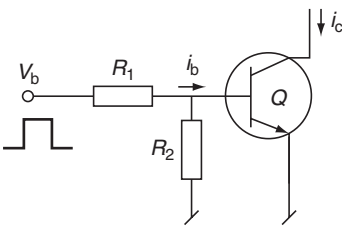


Figure 15.3 | Problem 15.3.

Solution

Circuit Description: This base drive circuit is the basic building block for many complex switching transistor drives. This is the simplest of all base drive circuits. For low-power and low-frequency applications, this base drive circuit is well suited. This circuit does not provide the starting surge base current i_{b+} during the turn-ON time. The resistor R_2 provides the discharge path for removing the base charges in the transistor Q. As a rule of thumb, it is found that $i_{bon} = 2\ (i_c/h_{FEsat})$ and $i_{R2} = i_{bon}$ gives satisfactory performance. The equations to determine the value of the resistors are

$$R_2 = \frac{V_{besat}}{i_{R2}} = \frac{V_{besat} \times h_{FEsat}}{2i_c} \tag{15.1}$$

$$R_1 = \frac{V_b - V_{besat}}{2i_{bon}} = \frac{(V_b - V_{besat}) \times h_{FEsat}}{4i_c} \tag{15.2}$$

The following is the functional specification for the above circuit:

Specifications

```
Vb=5; //volts pulse amplitude
fs=20e3;    //switching frequency in Hz
ic=1; //amps maximum during on time
```

Step 1: Generate the *design template* for the circuit. The template file is as follows.

```
//-------------------------------------------------

Q(1).type =
Q(1).PE_type =
Q(1).PQ_type =
Q(1).Tj =
Q(1).VCE =
Q(1).Vr =
Q(1).Pr =
Q(1).PA_type =
Q(1).Hp=

R(1).type =
R(1).PE_type =
R(1).PQ_type =
R(1).T =
R(1).S =
R(1).R =
R(1).Hp=

R(2).type =
R(2).PE_type =
R(2).PQ_type =
R(2).T =
R(2).S =
R(2).R =
```

Step 2: Enter the specifications and the design equations in the generated design template at the appropriate lines. This would actually be the functional design activity that the designer has to incorporate into the design template. The updated circuit-design file will now be as follows.

```
//------------SPECIFICATION OF THE CIRCUIT-------------

Vb =5;
fs =20e-3;
ic =1;
Vbe_sat=0.8;
hfe_sat=50;
Rat_pow_tr= 0.8;
Rat_pow_res= 0.25;
Vce_sat=0.2;
Vce_rated = 80;
T_amb =28;
R_thermal =187; // Thermal Resistance junction to Ambient
ib_on = 2*ic/hfe_sat;
//-------------------------------------------------

Q(1).type = 1;
Q(1).PE_type = 1;
```

```
Q(1).PQ_type = 4;
Q(1).Tj = T_amb + R_thermal*(Vce_sat*ic*0.5); //Ave. Conduction
  losses
Q(1).VCE = Vce_sat;
Q(1).Vr = Vce_rated;
Q(1).Pr = Rat_pow_tr;
Q(1).PA_type = 2;
Q(1).Hp=BJT(Q(1).type,Q(1).PE_type,Q(1).PQ_type,Q(1).Tj,Q(1).
  VCE,Q(1). Vr,Q(1).Pr,Q(1).PA_type);

R(1).type = 1;
R(1).PE_type = 1;
R(1).PQ_type = 6;
R(1).T = T_amb;
R(1).power =(Vb - Vbe_sat)*2*ib_on*0.5; // Avg power calculated from
  the switching waveform.
R(1).S = R(1).power /Rat_pow_res;
R(1).R = (Vb - Vbe_sat)/(2*ib_on);
R(1).Hp=resistor(R(1).type,R(1).PE_type,R(1).PQ_type,R(1).T,R(1).
  S,R(1).R);

R(2).type = 1;
R(2).PE_type = 1;
R(2).PQ_type = 6;
R(2).T = T_amb;
R(2).power= Vbe_sat*ib_on*0.5;
R(2).S = R(2).power /Rat_power;
R(2).R = (Vbe_sat)/ib_on;
R(2).Hp=resistor(R(2).type,R(2).PE_type,R(2).PQ_type,R(2).T,R(2).S,
  R(2).R);
```

Step 3: Execute the program and display the results. The result display will be as follows.

```
RELIABILITY PREDICTION FOR THE CIRCUIT
```

Part No.	type	PE_type	PQ_type	Tj	VCE	Vr	Pr	PA_type	Hp
Q1	1	1	4	46.7	0.2	80	0.8	2	0.19254
Part No.	type	PE_type	PQ_type	T	power	S	R	Hp	
R1	1	1	6	28	0.168	0.672	52.5	8.6894	
R2	1	1	6	28	0.016	0.064	20	2.8429	

```
CATEGORY-WISE HAZARD RATE
1.  BJT, FITs          =       0.19254
2.  Resistor, FITs     =       11.5323
TOTAL HAZARD RATE, FITs =    11.7249
MEAN TIME TO FAILURE (MTTF) =     9736.1778 years
```

This result is used as the basis for design iteration for derating and improving portions of the circuit having high failure rates.

15.11 Physics of Failure (PoF)

Currently, reliability prediction is based primarily on "Part Stress Analysis" or the co-variate approach. Here the failure rates are probabilistic models that have been developed using accelerated failure induction. These models are valid for large samples. However, they cannot be used for predicting the failure of individual components. If one has to focus on individual component failure, then the failure mechanisms of the component has to be well understood. This means a physics of failure (PoF) model has to be evolved for each individual component. Reliability prediction if based on PoF models would then be more accurate in predicting the failures of even the individual components.

A central feature of the PoF approach is that reliability modeling (time-to-failure modeling) is based on root cause failure processes or mechanisms. Failure mechanism models explicitly address the design, material, process and environment parameters causing the failures. In this respect, PoF focuses on the following three broad sections:

1. Identification of all possible failure mechanisms for the device.
2. Mathematical representation (model) of failure mechanisms as a function of time. In predicting the reliability of systems, some parts will have the PoF models available and some may not have the PoF models as yet available. In this context, how does one mix the co-variate approach and PoF approach such that one obtains failure rates that are more accurate than using only co-variate approach.

The shortcomings of traditional approach to design for reliability are becoming prominent as the manufacturing technology of electronic components improves. These design and reliability practices, based on stochastic models, discourage the use of newer, state-of-the-art, low-cost components as the use of newer components would result in lower life prediction because of less knowledge and confidence in them. The part stress methods for predicting the product life are therefore being discredited. More deterministic and scientific methods are required for assessing failure rates and identifying a validated model that considers all potential failure mechanisms. The PoF methodology enhances prediction of reliability by addressing the failure root cause by modeling failure mechanisms and the application environment. PoF approach enables prediction of failures within the operating lifetime of the product and the designer can then eliminate such failures by design techniques.

The "physics-of-failure" approach to reliability prediction has the following problems that prevent it being used currently:

1. All possible failure mechanisms are yet to be identified and therefore not available.
2. Different failure mechanisms result in different predicted failure times.

The most prominent device failure-rate prediction methodology is the US MIL-HDBK-217, which uses a co-variate model. The basic concept of MIL-HDBK-217 is to use historical part failure data to build an estimate of future system reliability. The handbook contains failure rate models for major electronic component

categories used in modern military systems. It contains two basic methods for performing a reliability prediction – the parts count method and the parts stress method. Both methods are based on empirical observations of historical data. The parts stress method is applied at the component level based on an analysis of component stresses by the system designer or reliability engineer. It uses a base failure rate for each component, which is a function of the applied stress or circuit complexity or package used. This failure rate is multiplied by a number of factors which account for environmental effects, quality and learning curve.

The parts count method is a special case of the parts stress method which assigns typical default values for the component stresses. Using either method, component failure rates are calculated for each component and summed for all components to estimate an overall system failure rate.

MIL-HDBK-217 has a few drawbacks. One is that it relies on field failure data. These are not easily available until well after a new technology has matured. Hence, it heavily penalizes new materials, structures and technology.

The PoF approach is a significantly more complex approach compared to the traditional MIL-HDBK-217 approach. This is because each and every potential failure mechanism must be analyzed to determine the MTTF. The failure mechanism with the shortest calculated life then becomes the weak link. Using this approach, the MTTF for the system is the same value as the calculated life of the weakest component making up the system. The PoF failure mechanism models are theoretically sound from the standpoint that they have been developed from existing principles of physics, chemistry, mechanics and materials science, however, they are not exhaustive.

EXAMPLE 15.4

Failure Mechanisms for Resistors

Two important failure mechanisms for the resistor component are described here.

***Electro-migration*:** Electro-migration is the movement of metal atoms of a metal line in the direction of the current flow through that metal line. This is a common failure mechanism in film resistors. As such, during electro-migration, metal atoms move from the starting end of the metal line and accumulate at the other end, forming voids at the entrance and hillocks at the exit of the metal line. Thus, electro-migration can result in open circuits (due to the voids) or line-to-line short circuits (due to the hillocks). The time to failure due to electro-migration is given as

$$t = CJ^{-n}e^{(E_a/KT)}$$

where C depends on metal properties; J the current density; n varies between 1 and 7 depending on material; E_a the activation energy; K the Boltzmann constant; T the temperature in K.

***Corrosion*:** This failure mechanism is observed in wire-wound resistors. Corrosion is metal degradation due to chemical or electrolytic reactions in the presence of moisture, contaminants and bias. The time to failure due to corrosion is given as

$$t = AV^{-n}H^{-m}e^{(E_a/KT)}$$

where V is the voltage across the resistor; H the relative humidity.

The problem with this approach is that all failure mechanisms have not been studied and/or modeled. Hence, the predicted life may be greater than the actual life if an unidentified failure mechanism were to be activated. Secondly, the method

is much more computationally complex than the co-variate approach. A third problem is that failure can occur in multiple ways, while PoF is deterministic. If PoF were to be believed, failure could occur only by that mechanism with the least predicted time to failure, but empirical observations show otherwise.

As PoF models are as yet in the nascent stage of development, only some of the PoF models for some of the failure mechanisms are available. Therefore, the best that can be done for now is to include deterministic knowledge of some of the PoF mechanisms into the co-variate models to improve the reliability prediction for such devices.

CONCLUDING REMARKS

The previous chapter discussed the methods for obtaining a reliability model for a given electronic system. In this chapter, the issue of predicting the reliability of an electronic circuit is addressed. The life specification of a system is in terms of the failure rate or mean time to failure. The main function of the reliability prediction is to estimate the failure rate of the designed system. The predicted failure rate is used to validate the circuit design in terms of the life specifications.

The reliability prediction has the added function of providing the necessary feedback to apply corrective action if the predicted failure rate does not meet the required life specifications. The corrective actions to improve the failure rate of the circuit is by (a) derating the critical parts; (b) reduction of stress by using snubbers; (c) modify the circuit and (d) use redundancy.

While reliability model is performed in a top-down manner from the system to the device level, the reliability prediction is performed in a bottom-up manner from the device to the system level. However, both need to be performed in order to quantify the reliability of a given system.

TUTORIAL EXERCISES

1. Study the user's manual of the REC toolbox (reliability for electronic circuits).
2. Study the sample circuits designed using the circuit template files generated by the REC toolbox that is included in the accompanying CD.
3. Observe the failure rate contribution of the various classes of the system.
4. Observe the failure rate contribution of the individual devices of the system.
5. Adjust the operating electrical and/or temperature stresses and re-execute to observe the changes in the failure rate of the system.
6. How does one estimate the number of gates within a digital IC?
7. How does one estimate the number of transistors in a linear IC?
8. Design a sample hybrid microcircuit and estimate its reliability using REC toolbox.

Term study task:
The additional information of the physics of failure model can be incorporated along with the co-variate model to form a hybrid reliability prediction method. Such a hybrid method will provide an estimate of the reliability that will not be as accurate as the deterministic physics of failure model, but will definitely be

better than the stochastic parts stress co-variate model. In this regard,

(a) Study the various failure mechanism from the physics of failure point of view.

(b) Evolve methods to combine the information of the physics of failure model with the co-variate prediction method.

FILL IN THE BLANKS

1. Parts count technique is used when the general _____ diagram is finalized but the _____ of the circuit components are not yet determined.

2. 1 FIT means one failure in _____ hours.

3. The components of the electronic circuits are modeled as _____ circuit.

4. Parts stress technique is used considering all operational _____.

5. Parts stress technique uses two extra co-variates as compared to the parts count technique that is _____ stress co-variate and _____ stress co-variate.

6. Most of the electronic products are designed for the _____ environment.

7. Reliability prediction involves the use of _____, _____ data models and failure rate equations.

8. In part stress analysis method, all the components are _____ and therefore from the reliability point of view the circuit is a _____ model.

DESCRIPTIVE QUESTIONS

1. What is reliability prediction?

2. Why does one need to perform reliability prediction?

3. What are the different methods of reliability prediction that are applicable at different stages of the product development?

4. Write short notes on: (a) similar equipment technique, (b) similar complexity technique and (c) similar function technique.

5. What is generic or base failure rate?

6. Discuss the environment co-variate.

7. Discuss the quality co-variate.

8. What are the different standards available used to determine the co-variates and the base failure rates?

9. The part stress technique is very much similar to the part count technique except that the thermal and electrical stresses are evaluated for each part.

10. What are the standard environments defined by the MIL-STD?

11. Explain the steps in the reliability prediction process.

12. Discuss the component naming convention.

13. Discuss the component object notation.

14. What is physics of failure model for components?

15. What are the advantages and limitations of physics of failure model?

16. Why is the physics of failure model not as popular as the co-variate model?

PROBLEMS

Use the reliability toolbox provided in the accompanying CD to estimate the reliability of the following circuits (Figures 15.4–15.6). Find out the class of components that contribute most to the failure rate of the system. Also identify the critical components of the system in terms of failure rate.

1. The specifications are as follows:
- **a.** Output voltage: 15 V DC
- **b.** Output current: 50–500 mA
- **c.** Input AC voltage: 230 V_{rms} ± 20%; 50 Hz

2. The specifications are as follows:
- **a.** Input voltage: V_{in} = 15 V
- **b.** Output voltage: V_o = 22 V
- **c.** Output current: $I_o = V_o/R_o$ = 100 mA
- **d.** Switching frequency: fs = 22.22 kHz
- **e.** Ripple in output voltage: 14 mV

3. The specifications are as follows:
- **a.** Output voltage: V_o = 5 V DC
- **b.** Output current: I_o = 10–250 mA
- **c.** Input voltage: V_{in} = 15–40 V DC

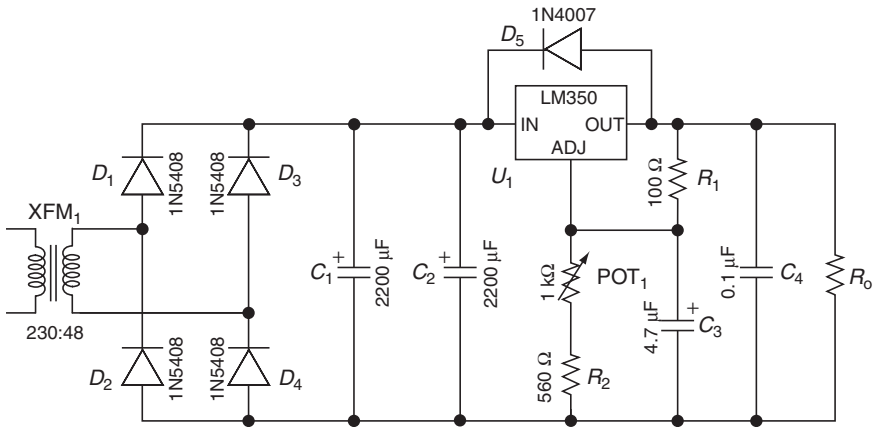


Figure 15.4 | Circuit 1.

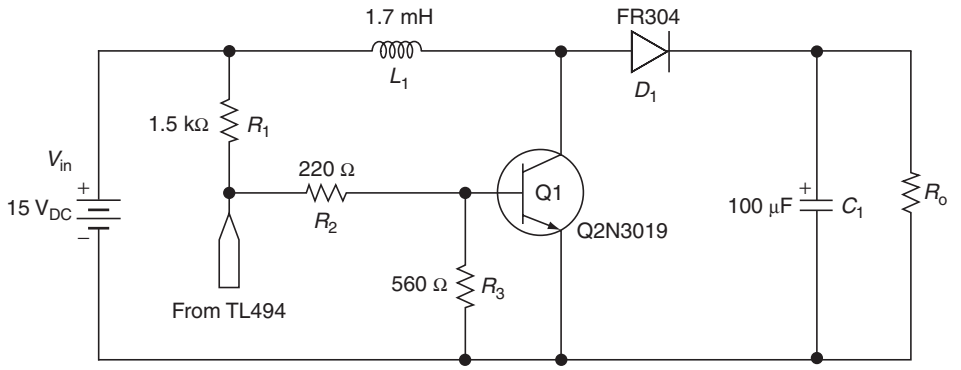


Figure 15.5 | Circuit 2.

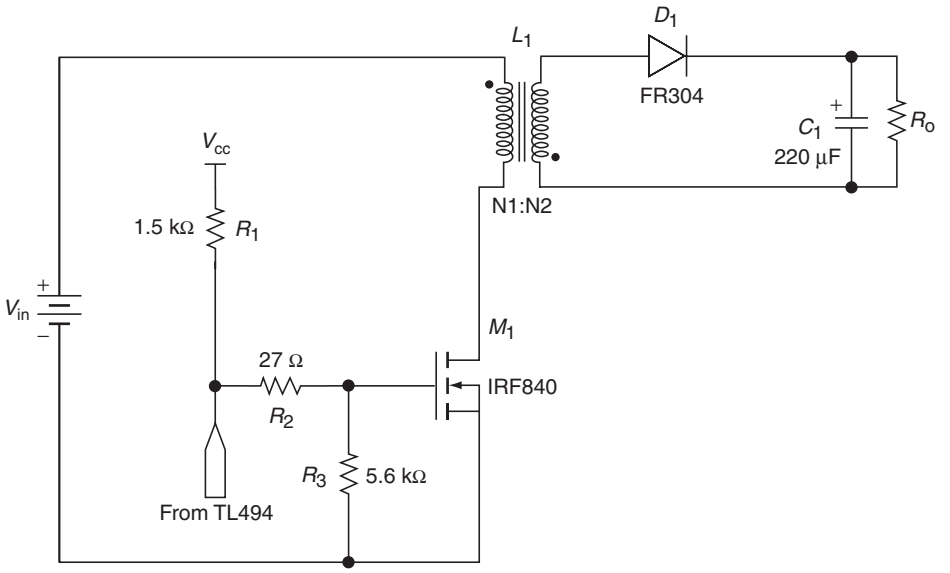


Figure 15.6 | Circuit 3.

d. Maximum peak-to-peak ripple in output voltage: $V_r = 0.05$ V
e. Switching period = 45 μs
f. Operating core flux density: $B_m = 0.25$ T
g. The control power supply voltage: $V_{cc} = 15$ V

ANSWERS

Fill in the Blanks

1. circuit; values
2. 10^9
3. series
4. stresses
5. temperature; electrical
6. ground benign
7. nomo-graphs; curve-fit
8. essential; series

Appendix I

Area Product for the Saturable-Core Transformer

Refer to Chapter 7 for a discussion on the area product approach to design of the transformers.

$$K_w A_w = N_{p1} a_{p1} + N_{p2} a_{p2} + N_{b1} a_{b1} + N_{b2} a_{b2} + N_s a_s \tag{A1.1}$$

where a_i is the wire cross-section area.

$$a_{p1} = \frac{i_{p1}}{J} = a_{p2}$$

$$a_{b1} = \frac{i_{b1}}{J} = a_{b2}$$

$$a_s = \frac{i_s}{J}$$

$$E_{p1} = N_{p1} \frac{d\phi}{dt} = N_{p1} A_c \frac{2B_m}{T_s / 2} = 4 N_{p1} A_c B_m f_s \tag{A1.2}$$

$$N_{p1} = N_{p2} = \frac{E_{p1}}{4 A_c B_m f_s} \tag{A1.3}$$

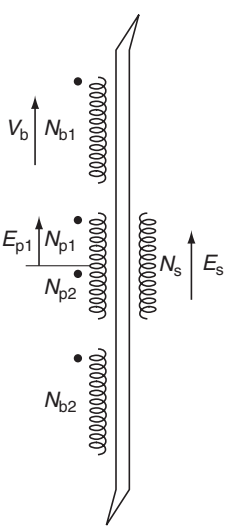


Figure A1 | Saturable-core transformer.

$$N_s = \frac{E_s}{4A_c B_m f_s} \tag{A1.4}$$

$$N_{b1} = N_{b2} = \frac{V_b}{4A_c B_m f_s} \tag{A1.5}$$

$$K_w A_w J = \frac{2E_{p1} i_{p1}}{4A_c B_m f_s} + \frac{2V_b i_{b1}}{4A_c B_m f_s} + \frac{E_s i_s}{4A_c B_m f_s} \tag{A1.6}$$

Simplifying Eq. (A1.6) we get

$$K_w A_w J = \frac{1}{4A_c B_m f_s}\left(2V_{in}\frac{I_{in}}{\sqrt{2}} + 2n_b V_{in}\frac{I_{in}}{h_{FE}\sqrt{2}} + E_s i_s\right) \tag{A1.7}$$

where V_{in} is the DC-link voltage for push–pull; I_{in} the current through V_{in} connected to center-tap of push–pull transformer.

$$n_b = \frac{V_b}{V_{in}} = \frac{N_{b2}}{N_{p2}} = \frac{N_{b1}}{N_{p1}}$$

$$i_{b1} = \frac{1}{\sqrt{2}}\left(\frac{I_{in}}{h_{FEmin}}\right)$$

Simplifying Eq. (A1.7), one obtains

$$K_w A_w J = \frac{1}{4A_c B_m f_s}\left(\sqrt{2}\cdot P_{in} + \sqrt{2}\cdot\frac{n_b}{h_{FEmin}}P_{in} + P_o\right) \tag{A1.8}$$

where $P_o = E_s i_s$. From Eq. (A1.8), the area product for the saturable-core transformer of Figure A1 is given by

$$A_p = \frac{P_o}{4K_w J B_m f_s}\left(\frac{\sqrt{2}}{\eta} + \frac{\sqrt{2}\cdot n_b}{\eta\cdot h_{FEmin}} + 1\right) \tag{A1.9}$$

Appendix II

Properties of Few Ferrite Cores

Cores without air gap	Mean length per turn (mm)	Mean magnetic length l_m (mm)	Core cross-section area A_c (mm^2)	Window area, A_w (mm^2)	Area product A_p (mm^4)
Pot Cores					
P18/11	35.6	26	43	27	1161
P26/16	52	37.5	94	53	4982
P30/19	60	45.2	136	75	10200
P36/22	73	53.2	201	101	20301
P42/29	86	68.6	264	181	47784
P66/56	130	123	715	518	370370
EE Cores					
E20/10/5	38	42.8	31	47.8	1481
E25/9/6	51.2	48.8	40	78	3120
E25/13/7	52	57.5	55	87	4785
E30/15/7	56	66.9	59.7	119	7104.3
E36/18/11	70.6	78	131	141	18471
E42/21/9	77.6	108.5	107	256	27392
E42/21/15	93	97.2	182	256	46592
E42/21/20	99	98	235	256	60160
E65/32/13	150	146.3	266	537	142842
UU Cores					
UU 15	44	48	32	59	1888
UU 21	55	68	55	101	5555
UU 23	64	74	61	136	8296
UU 60	183	184	196	1165	228340
UU 100	29.3	308	645	2914	1879530
Toroids					
T 10	12.8	23.55	6.2	19.6	121.52
T 12	19.2	30.4	12	44.2	530.4
T 16	24.2	38.7	20	78.5	1570
T 20	25.2	47.3	22	95	2090
T 27	34.1	65.94	42	165.1	6934.2
T 32	39.6	73	61	165.1	10071.1
T 45	54.7	114.5	93	615.7	57260.1

Standard Types for Transformers and Choke Laminations

Type No.	A_c (mm^2)	A_w (mm^2)	A_p (mm^4)
17	161	120	19320
12A	252	188.4	47476.8
21	252	332	83664
10	252	443	111636
10A	252	443	111636
1	278.9	654.3	182484.27
74	306.3	228	698364
23	363	272	98736
11	363	726	263538
11A	363	907	329241
2	363	1089	395307
30	400	300	120000
31	494	371	183274
45	494	371	183274
15	645	484	312180
44	645	484	312180
14	645	655	422475
4	645	1585	1022325
33	784	588	460992
3	1008	754	760032
13	1008	1411	1422288
4A	1112	1026	1140912
16	1450	1089	1579050
5	1450	1273	1845850
6	1450	1935	2805750
7	2580	1895	4889100
8	2580	5123	13217340
34	250	668	167000
9	440	910	400400
9A	490	790	387100
4AX	560	1307	731920
75	650	1526	991900
35A	1450	3934	5704300
8B	5810	6552	38067120
100	10320	11610	119815200

Appendix III

Wire Size Table

SWG	Diameter with enamel (mm)	Area of bare conductor (mm^2)	*R*/km @20°C Ω	Weight (kg/km)
45	0.086	0.003973	4,340	0.0369
44	0.097	0.005189	3,323	0.0481
43	0.109	0.006567	2,626	0.061
42	0.119	0.008107	2,127	0.075
41	0.132	0.009810	1,758	0.0908
40	0.142	0.011675	1,477	0.1079
39	0.152	0.013700	1,258	0.1262
38	0.175	0.018240	945.2	0.1679
37	0.198	0.023430	735.9	0.2202
36	0.218	0.029270	589.1	0.2686
35	0.241	0.035750	482.2	0.3281
34	0.264	0.042890	402	0.3932
33	0.287	0.050670	340.3	0.465
32	0.307	0.059100	291.7	0.5408
31	0.33	0.06818	252.9	0.6245
30	0.351	0.07791	221.3	0.7121
29	0.384	0.09372	184	0.8559
28	0.417	0.11100	155.3	1.014
27	0.462	0.13630	126.5	1.245
26	0.505	0.16420	105	1.499
25	0.561	0.20270	85.1	1.851
24	0.612	0.24520	70.3	2.233
23	0.665	0.29190	59.1	2.655
22	0.77	0.39730	43.4	3.607
21	0.874	0.51890	33.2	4.702
20	0.978	0.65670	26.3	5.939
19	1.082	0.81070	21.3	7.324

SWG	Diameter with enamel (mm)	Area of bare conductor (mm^2)	*R*/km @20°C Ω	Weight (kg/km)
18	1.293	1.16700	14.8	10.537
17	1.501	1.589	10.8	14.313
16	1.709	2.075	8.3	18.678
15	1.92	2.627	6.6	23.64
14	2.129	3.243	5.3	29.15
13	2.441	4.289	4	38.56
12	2.756	5.48	3.1	49.22
11	3.068	6.818	2.5	61
10	3.383	8.302	2.1	74
9	3.8	10.51	1.6	94
8	4.219	12.97	1.3	116

Appendix IV

Laplace Transforms and z-Transforms of Few Time Domain Signals

Time function	Laplace transform	z-transform
$\delta(t)$	1	1
$\delta(t-kT)$	e^{-kTs}	z^{-k}
$u(t)$ – unit step	$\frac{1}{s}$	$\frac{z}{z-1}$
T	$\frac{1}{s^2}$	$\frac{Tz}{(z-1)^2}$
$\frac{t^2}{2}$	$\frac{1}{s^3}$	$\frac{T^2 z(z+1)}{2(z-1)^3}$
e^{-at}	$\frac{1}{s+a}$	$\frac{z}{z-e^{-aT}}$
te^{-at}	$\frac{1}{(s+a)^2}$	$\frac{Tze^{-aT}}{(z-e^{-aT})^2}$
$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$	$\frac{z \sin \omega T}{z^2-2z\cos \omega T+1}$
$\cos \omega t$	$\frac{s}{s^2+\omega^2}$	$\frac{z^2-z\cos \omega T}{z^2-2z\cos \omega T+1}$
$e^{-at}f(t)$	$F(s+a)$	$f(e^{aT}z)$

Appendix V

Induction Motor Specifications

3 phase, 400 V rms, 50 Hz, 4 pole machine ($p = 2$)

$R_s = 0.19\ \Omega$

$R_r = 0.125\ \Omega$

$M = 36.9$ mH

$L_{ss} = 38.51$ mH

$L_{rr} = 37.56$ mH

Nominal drive torque = 98 Nm

$J = 0.1$ Nms^2/rad

$B = 0.01$ Nms/rad

Appendix VI

Component Parameter Description

BJT Q

```
-----------------------------------------------------------------------
                    SELECT THE BJT TYPE

         1. Transistors,low frequency,bipolar
         2. Transistors,low noise,high frequency,bipolar
         3. Transistors,high power,high frequency,bipolar
-----------------------------------------------------------------------
-----------------------------------------------------------------------
   Type 1 : Transistors,low frequency,bipolar
   Hp=BJT(type,PE_type,PQ_type,Tj,VCE,Vr,Pr,PA_type)
     PE_type: 1  2  3  4  5  6    7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
     PQ_type:                 1       2      3      4        5
   PQ_type=Quality factor: JANTXV   JANTX   JAN   LOWER   PLASTIC
   Tj=Junction temperature(C)
   VCE=Operating collector to emitter voltage
   Vr=Rated collector to emitter voltage,base open
   Pr=Rated power in watt
                PA_type=            1                2
   PA=Application factor=    linearamplifcation   switching
-----------------------------------------------------------------------
-----------------------------------------------------------------------
     Type 2 :  Transistors,low noise,high frequency,bipolar
     Hp=BJT(type,PE_type,PQ_type,Tj,VCE,Vr,Pr)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
     PQ_type:                   1       2      3      4
     PQ_type=Quality factor; JANTXV   JANTX   JAN   LOWER
     Tj=Junction temperature(C)
     VCE=Operating collector to emitter voltage
     Vr=Rated collector to emitter voltage,base open
     Pr=Rated power in watt
-----------------------------------------------------------------------
```

```
-----------------------------------------------------------------------
     Type 3 :  Transistors,high power,high frequency,bipolar
Hp=BJT(type,PE_type,PQ_type,Tj,VCE,Vr,PA_type,DF,P,F,PT_type,PM_type)
     PE_type: 1   2   3   4   5   6    7   8    9    10   11  12  13  14
Env. factor : GB  GF  GM  NS  NU  AIC  AIF AUC  AUF  ARW  SF  MF  ML  CL
        PQ_type:                    1         2       3       4
     PQ_type=Quality factor:     JANTXV    JANTX    JAN    LOWER
     Tj=Junction temperature(C)
     VCE=Operating collector to emitter voltage
     Vr=Rated collector-emitter breakdown voltage with base shorted to
        emitter(volts)
     PA_type =1 For CW application
              2 For pulsed application
     DF=Duty factor=[1 5 10 15 20 25 30]
     P=Output power(W)
     F=Frequency(GHz)
     PT_type=1 For gold metallization
             2 For aluminum metallization
     PM_type=                          1               2         3
     PM=Matching network factor=Input and output     Input     None
-----------------------------------------------------------------------
```

Capacitor C

```
function Hp =capacitor(type,PE_type,PQ_type,T,S,C,Rc,Vc),
SYNTAX:
Hp =capacitor(type,PE_type,PQ_type,T,S,C,Rc,Vc)
```

type	Capacitor style	Spec MIL-C-	Description
1.	CP	25	Fixed, paper DC (hermetic seal in metal case)
2.	CA	12889	Bypass, Radio interference rejection paper, AC/DC (hermetic,metal case)
3.	CZ,CZR	11693	Feed through, Radio interference reduction, AC/DC (hermetic, metal case) Est.R and non-Est. R
4.	CQ,CQR	19978	Fixed, plastic/paper plastic (hermetically sealed in metal/

			ceramic/glass cases) Est.R and non-Est.R
5.	CH	18312	Fixed, metallized (paper, paper plastic plastic flm), DC (hermetically sealed in metal cases)
6.	CHR	39022	Fixed, metallized (paper, paper plastic plastic flm)
7.	CFR	55514	Fixed, plastic (or metallized plastic) DC in non-metal cases
8.	CRH	83421	Fixed, supermetallised plastic flm (DC, AC or DC and AC) hermetic sealed in metal cases, Est. R
9.	CM	5	Fixed, mica dielectric
10.	CMR	39001	Fixed, mica dielectric, Est.R
11.	CB	10950	Fixed, mica dielectric, button style
12.	CY	11272	Fixed, glass dielectric
13.	CYR	23269	Fixed, glass dielectric, Est.R
14.	CK	11015	Fixed, ceramic, general purpose
15.	CKR	39014	Fixed, ceramic, general purpose, Est.R
16.	CC,CCR	20	Fixed, ceramic (temperature compensating)Est.R and non-Est.R
17.	CDR	55681	Chip, Multiple layer, fxed, ceramic dielectric, Est. R
18.	CSR	39003	Fixed, electrolytic (solid electrolyte), Tantalum, Est.R
19.	CWR	55365	Fixed, electrolytic (Tantalum) chip, Est.R
20.	CL	3965	Fixed, electrolytic (non-solid electrolyte)Tantalum
21.	CLR	39006	Fixed, electrolytic (non-solid electrolyte)Tantalum, Est.R
22.	CRL	83500	Fixed, electrolytic (non-solid electrolyte)Tantalum cathode
23.	CU,CUR	39018	Fixed, electrolytic (Aluminum oxide Est.R and non-Est.R
24.	CE	62	Fixed, electrolytic (DC, Aluminum, Dry electrolyte, polarized)

```
  25.     CV          81       Variable, ceramic (Trimmer)
  26.     PC          14409    Variable, (piston type, tubular
                                trimmer)
  27.     CT          92       Variable, air dielectric(Trimmer)
  28.     CG          23183    Fixed or variable, vacuum dielectric
-----------------------------------------------------------------------
     PE_type: 1   2   3  4  5   6   7   8   9   10  11  12  13  14
Env. factor : GB  GF  GM NS NU  AIC AIF AUC AUF ARW SF  MF  ML  CL
     PQ_type:    1   2   3   4  5  6  7        8                9
 Quality factor: D   C  S,B  R  P  M  L  MIL-C-xx,non-Est.R.  Lower
 T=Capacitor ambient temperature(C)
 S=Ratio of operating voltage to rated voltage
 C=Capacitance in micro farad
 Rc=Effective resistance between capacitor and power supply for
    Tantalum CSR style capacitors only
 Vc=Voltage applied to capacitor for Tantalum CSR style capacitors
    only
-----------------------------------------------------------------------
```

Connections SC

```
function Hp=connections(type,PE_type,PQ_type,N1,N2,P),
-----------------------------------------------------------------------
  type:
    1. Hand solder, w/o wrapping
    2. Hand solder, w/ wrapping
    3. Crimp
    4. Weld
    5. Solderless wrap
    6. Clip termination
    7. Refow solder
    8. Spring contact
    9. Terminal block

       INTERCONNECTIONS
   10. Printed wiring assembly/PCBs with PTHs
   11. Discrete wiring with electroless deposited PTH (<=2 levels of
       circuitry)
-----------------------------------------------------------------------
     PE_type:  1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor :  GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
           PQ_type :      1           2
     Quality factor:   MIL-SPEC      Lower
```

```
  N1 : Number of PTHs by automated techniques (Quantity of wave
       infrared (IR) or vapour phase soldered functional PTHs)
  N2 : Number of hand soldered PTHs
  P  : Number of circuit planes or layers (2<= P <= 18)
-----------------------------------------------------------------------
```

Connectors J

```
function Hp=connectors(type,PE_type,PQ_type,a1,a2,a3,a4)
-------------------------------------------------------
  type: type of connector

    1. Circular/Cylindrical
    2. Card Edge (PCB)
    3. Hexagonal
    4. Rack and panel
    5. Rectangular
    6. RF Coaxial (low power)
    7. RF Coaxial (high power)
    8. Telephone
    9. Power
   10. Triaxial

       IC SOCKETS
   11. Dual-in-line package
   12. Single-in-line package
   13. Chip carrier
   14. Pin Grid Array
   15. Relay socket
   16. Transistor socket
   17. Electron tube, CRT socket
-----------------------------------------------------------------------

     PE_type:  1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor :  GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL

           PQ_type :       1            2
     Quality factor:    MIL-SPEC      Lower

  G : Contact Gauge :   1. 32 Gauge contact
                        2. 30 Gauge contact
                        3. 28 Gauge contact
                        4. 24 Gauge contact
                        5. 22 Gauge contact
                        6. 20 Gauge contact
                        7. 18 Gauge contact
                        8. 16 Gauge contact
                        9. 12 Gauge contact
```

```
  i  : Amperes per contact
  TA : connector ambient (C)
  C=Mating/unmating cycles(per 1000 hours)
  N=Number of active contacts (for IC socket connectors)
-----------------------------------------------------------------
```

Diode D

```
function Hp=diode(type,D_type,PE_type,PQ_type,Tj,a1,a2,a3),
-----------------------------------------------------------------
  SELECT THE DIODE TYPE

    1.Low frequency diode
    2.High frequency(microwave,RF) diode
-----------------------------------------------------------------
-----------------------------------------------------------------
       Type 1: Low frequency
       Hp=diode(type,D_type,PE_type,PQ_type,Tj,PC,Vo,Vr)
       D_type
           1 General purpose analog
           2 Switching
           3 Power rectifer,fast recovery
           4 Power rectifer,schottky power diode
           5 Power rectifer with high voltage stacks
           6 Transient suppressor/varistor
           7 Current regulator
           8 Voltage regulator and voltage reference(avalanche and
             zener)
      PE_type: 1   2   3   4  5  6    7    8    9    10   11 12 13 14
Env. factor : GB  GF  GM  NS NU AIC  AIF  AUC  AUF  ARW  SF MF ML CL
              PQ_type:        1      2      3      4         5
       Quality factor:     JANTXV  JANTX   JAN   LOWER    PLASTIC
       Tj=Operating junction temperture(C)
       PC=1 For metallurgical bonded
         =2 For nonmetallurgical bonded & spring loaded constants
       Vo= Operating peak inverse voltage (PIV), volts
       Vr= Rated peak inverse voltage (PIV), volts
-----------------------------------------------------------------
-----------------------------------------------------------------
       Type 2 : high frequency(microwave,RF)
       Hp=diode(type,D_type,PE_type,PQ_type,Tj,PA_type,Pr)
```

```
      D_type
          1 Si IMPATT(<=35GHz)
          2 Gunn/bulk effect
          3 Tunnel and back(including mixer,detectors)
          4 Pin
          5 Schottky barrier(including detectors)and point contact
            (200MHz<=frequency<=35GHz)|
          6 Varactor and step recovery
     PE_type: 1   2   3   4  5  6    7    8    9    10   11 12 13 14
Env. factor : GB  GF  GM  NS NU AIC  AIF  AUC  AUF  ARW  SF MF ML CL
            PQ_type:     1         2       3       4         5
     Quality factor:   JANTXV    JANTX    JAN    LOWER    PLASTIC
     Tj=Operating junction temperature (C)
     Application factor - PA_type: 1 For varactor, voltage control,
                                   2 For varactor multiplier,
                                   3 All other diode=1.0
     Pr=Rated power, watts
-----------------------------------------------------------------------
```

Electronic Filters EF

```
function Hp=electronicflters(PE_type,PQ_type,Hb_type),
-----------------------------------------------------------------------
   Hp=electronicflters(PE_type,PQ_type,Hb_type)
     PE_type: 1  2   3  4  5  6   7   8   9   10  11 12 13 14
Env. factor : GB GF  GM NS NU AIC AIF AUC AUF ARW SF MF ML CL
            PQ=Quality factor:  PQ_type :  1 For SPEC
                                           2 For lower
        Hb=Base failure rate;
 Hb_type =1 For MlL-F-l5733,Ceramic-Ferrite Construction
            (Styles FL 10-16,22,24,30-32,34,35,38,41-43,45,47-50,
             61-65,70, 81-93,95,96)
          2 For MlL-F-l5733, Discrete LC Components, (Styles FL 37,
            53, 74|
          3 For MlL-F-l8327, Discrete LC Components (Composition 1)
          4 For MIL-F-18327, Discrete LC and Crystal Components
1.9.7. Fuses-FU
function Hp=fuses(PE_type),
-----------------------------------------------------------------------
     Hp=fuses(PE_type)
     PE_type: 1   2   3   4  5  6    7    8    9    10   11 12 13 14
Env. factor : GB  GF  GM  NS NU AIC  AIF  AUC  AUF  ARW  SF MF ML CL
```

Inductance L

```
function Hp=inductance(type,PE_type,PQ_type,THS),
------------------------------------------------------------  COILS:-
     Hp=inductance(type,PE_type,PQ_type,THS)
     type:     1. Fixed inductor or Choke
               2. Variable inductor
     PE_type: 1  2  3  4  5  6   7   8   9   10   11 12 13 14
Env. factor : GB GF GM NS NU AIC AIF AUC AUF ARW  SF MF ML CL
                    PQ_type:   1    2    3     4     5          6
     PQ_type=Quality factor:   S    R    P     M   MIL-SPEC    Lower
     THS=Hot spot temperature(C)
```

Lamps LA

```
function Hp=lamps(PE_type,PA_type,PU,Vr),
-----------------------------------------------------------------------
     Hp=lamps(PE_type,PA_type,PU,Vr)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11 12 13 14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF MF ML CL
     PA:Application factor:PA_type:1 For alternating current
                                   :2 For direct current
     PU=Utilization factor:Utilization(illuminate hrs/eqpt operate
                           hrs)
     Vr=Rated voltage(volts)
```

Laser LAS

```
function Hp=Laser(type,PE_type,a1,a2,a3,a4,a5,a6,a7,a8,a9,a10,a11,a12)
-----------------------------------------------------------------------
           SELECT THE LASER TYPE
           1 Helium and argon
           2 Carbon dioxide,sealed
           3 Carbon dioxide,fowing
           4 Solid state,ND:YAG and ruby rod
-----------------------------------------------------------------------
     Type 1 : helium and argon
     Hp=Laser(type,PE_type,HM_type,HC_type)
     PE_type: 1   2   3   4  5  6    7    8    9    10   11 12 13 14
Env. factor : GB  GF  GM  NS NU AIC  AIF  AUC  AUF  ARW  SF MF ML CL
     HM=Lasing media failure rate:- HM_type:    1        2       3
                                      type:  He/Ne    He/Cd   Argon
     HC=Coupling failure rate:- HC_type:1 For helium
                                        2 For argon
```

```
------------------------------------------------------------------
     Type 2 : carbon dioxide,sealed
     Hp=Laser(type,PE_type,I,POS,OF,BVI)
     PE_type:  1  2  3  4  5  6    7    8    9   10   11  12  13  14
Env. factor : GB  GF GM NS NU AIC  AIF  AUC  AUF ARW  SF  MF  ML  CL
     I=Tube current(mA)  10<=I<=150
     POS=Optical surface factor:POS=1 For 1 active optical surface
                                =2 For 2 active optical surface
     OF=CO2 overfll percent(0 to 50 )
     BVI=Percent of ballast volumetric increase ( 0 to 200 )
------------------------------------------------------------------
     Type 3 : carbon dioxide,fowing
     Hp=Laser(type,PE_type,POS,P)
     PE_type:  1  2  3  4  5  6    7    8    9   10   11  12  13  14
Env. factor : GB  GF GM NS NU AIC  AIF  AUC  AUF ARW  SF  MF  ML  CL
     POS=Optical surface factor:POS=1 For 1 active optical surface
                                =2 For 2 active optical surfacev
     P=Avg power output in KW.  0.01<=P<=1.0
------------------------------------------------------------------
     Type 4: solid state,ND:YAG and ruby rod
     Hpump_type=Pump Pulse Failure Rate:-1 For xenon Flashlamps.
it is the failure rate contribution of the xenon fashlamp or
fashtube.
The fash lamps evaluated herein are linear types.
                                        2 For krypton fashlamps.
it is the failure rate contribution Of the krypton fashlamp or
fashtube.
The fashlamps evaluated herein are the continuouswave(CW) type.they
are approximatety 7mm in diameter and 5 to 6 inches long.|
------------------------------------------------------------------
     Type=4:- Hpump_type : 1 for xenon Flashlamps
Hp=Laser(type,PE_type,POS,PC_type,Hpump_type,Pcool,L,HM_type,PPS1,
Ej,d,t,F,PPS2)
     PE_type: 1  2  3  4  5  6    7    8    9   10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF ARW  SF  MF  ML  CL
     POS=Optical surface factor:POS=1 For 1 active optical surface
                                =2 For 2 active optical surface
     PC_type=Coupling cleanliness factor type:-
PC_type
 :1 For Cleanliness level:regorous cleanliness procedures and
    trained maintenance personnel.bellows provided over optical train
```

```
 :2 For Cleanliness level:Mlnimal precautions during opening,
    maintenance, repair, and testing.bellows provided over optical
    train
 :3 For Cleanliness level:Mlnimal precautions during opening,
    maintenance, repair, and testing.no bellows provided over optical
    train
     Hpump_type=1
Pcool=Is the cooling factor due to various cooling media immediately
surrounding the fashlamp or fashtube,Pcool=1.0 for any air or inef gas
cooling,Pcool =0.1 for all liquid cooled designs.
Default value:Pcool=0.1,lquid cooled.
     L=Is the fashlamp or fashtube arc length in inches.
           Default value: L = 2.
     HM_type:Media failure rate:-1 For  ND:YAG
                                 2 For  ruby
PPS1=It is the repetition pulse rate in pulses per second.
 Typical values range between 1 & 20 pulses per second.
Ej=It is the fashlamp or fashtube input energy per pulse, in joules.
 Its value is determined from the actual or design input energy. For
 values less than 30 joules.use Ej=30 Default value:Ej=40.
d=It is the fashlamp or fashtube inside diameter, in millimeters.
 Default value: d = 4
t=It is the truncated pulse width in microseconds.
 Use t -100 microseconds.for any truncatd pulse width exceeding 100
 microseconds. For shotier duration pulses, pulse width is to be
 measured at 10 percent of the maximum current amplitude.
     Default value: t = 100
     Only if HM_type=2 enter the value of F & PPS2.
F=Is the energy density in Joules per cm^2/pulse over the
 cross-sectional area of the laser beam.which is nominally
 equivalent to the cross-sectional area of the laser rod.
      PPS2=Is the number of pulses per second.
-------------------------------------------------------
     Type=4:- Hpump_type 2 :for krypton fashlamps
     Hp=Laser(type,PE_type,POS,PC_type,Hpump_type,Pcool,L,HM_
type,P,F,PPS2)
     PE_type: 1  2  3   4  5  6    7    8  9    10   11  12  13  14
Env. factor : GB GF GM  NS NU AIC  AIF AUC AUF  ARW  SF  MF  ML  CL
     POS=Optical surface factor:POS=1 For 1 active optical surface
                                   =2 For 2 active optical surface
     PC_type=Coupling cleanliness factor type
```

```
PC_type
 :1 For Cleanliness level:regorous cleanliness procedures and
    trained maintenance personnel.bellows provided over optical
    train
 :2 For Cleanliness level:Mlnimal precautions during opening,
    maintenance, repair, and testing.bellows provided over optical
    train
 :3 For Cleanliness level:Mlnimal precautions during opening,
    maintenance, repair, and testing.no bellows provided over
    optical train
     Hpump_type=2
     Pcool=Is the cooling factor due to various cooling media
     immediately surrounding the fashlamp or fashtube.
           Pcool=1 for any air or inert gas cooling.
           Pcool=0.1 for all liquid designs.
           Defautt value:- Pcool=0.1,liquidcooled.
     L=Is the fashlamp or fashtube arc length in inches.
        Defautt value: L =2
     HM_type:Media failure rate:-1 For  ND:YAG
                                2 For  ruby
     P=Is the average inputpower in Kilowatts.Default
       value: P =4
     only if HM_type=2 Enter the value of F & PPS2.
     F=Is the energy density in Joules per cm^2/pulse over the
        cross-sectional area of the laser beam.which is
        nominally equivalent to the cross-sectional area of the
        laser rod.
      PPS2=Is the number of pulses per second.
```

Meters Panel MEP

```
function Hp=meterspanel(PE_type,PQ_type,PF_type,PA_type),
----------------------------------------------------------------
     Hp=meterspanel(PE_type,PQ_type,PF_type,PA_type)
     PE_type: 1  2  3  4  5   6    7    8    9    10  11  12  13  14
Env. factor : GB GF GM NS NU  AIC  AIF  AUC  AUF  ARW SF  MF  ML  CL
     PQ=Quality factor:PQ_type : 1 For MIL-M-10304
                               : 2 For lower
     PF=Function factor:PF_type : 1 For ammeter
                                : 2 For voltmeter
                                : 3 Other
     PA:Application factor:PA_type:1 For direct current
                                   2 For alternating current
```

Microcircuits: Bubble Memories

```
function Hp=microcircuits_bubblememories(PE_type,PQ_type,y,pkg,Np,
Tc,NC,N1,N2,D1,D2,RW)
------------------------- ----------------------------------------
     Hp=microcircuits_bubblememories(PE_type,PQ_type,y,pkg,Np,Tc,NC,
     N1,N2,D1,D2,RW)
     PE_type: 1  2  3  4  5   6   7    8   9   10   11  12  13  14
Env. factor : GB GF GM NS NU  AIC AIF  AUC AUF ARW  SF  MF  ML  CL
     Quality class:
     PQ_type= 1 For class S categories:a.procured in full accordance
                with MIL-M-38510.class S requirements.
              2 For class B categories: a.procured in full
                accordance with MIL-M-38510.class B requirements.
              3 For class B-1 category:-procured to a in
                production MIL drawing. DESC drawing or the
                government approved documentation.
              4 Other commercial or unknown screening levels
     y=Number of years the device has been in production
     pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                         gnd array(PGA)1.SMT(leaded and nonleaded)
                       2 For DIPs with glass
                       3 For fatpacks with axial leads on 50 Mil
                         centers
                       4 For cans
                       5 For nonhermetic DIPs,PGA,SMT(Leaded and
                         nonleaded)
     Np=Number of functional pins
     Tc=Case temperature(C)
     NC=Number of bubble chips per packaged device
     N1=Number of dissipative elements on  a
         chip(gates,detectors,generators..etc)N1<=1000
     N2=Number of bits,N2<=9*10^6
     D1=Avg device data rate
     D2=Mfg. max. rated data rate  D=(D1/D2) <=1
     RW=No. of reads per write
```

Microcircuits Digital

```
function Hp=microcircuits_digital_ic(PE_type,PQ_type,y,pkg,Np,
family,N,Tj),
-----------------------------------------------------------------------
     Hp=microcircuits_digital_ic(PE_type,PQ_type,y,pkg,Np,family,
        N,Tj)
```

```
     PE_type: 1  2  3  4  5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF GM NS NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
     Quality class:
     PQ_type= 1 For class S categories:a.procured in full accordance
                with MIL-M-38510.class S requirements.
              2 For class B categories:a.procured in full accordance
                with MIL-M-38510.class B requirements.
              3 For class B-1 category:-procured to a in production
                MIL drawing. DESC drawing or dher government
                approved documentation.
              4 Other commercial or unknown screening levels|
     y=Number of years the device has been in production
   pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                       gnd array(PGA)1.SMT(leaded and nonleaded)
                     2 For DIPs with glass
                     3 For fatpacks with axial leads on 50 Mil
                       centers
                     4 For cans
                     5 For nonhermetic DIPs,PGA,SMT(Leaded and
                       nonleaded)
     Np=Number of functional pins
     family : 1 For TTL,STTL,CML,HTTL,FTTL,DTL,FCL,ASTTL
              2 For BiCMOS,LSTTL,LTTL,ALSTTL
              3 For III,ISL
              4 For digital MOS,CMOS
     N=Number of gates in the IC
     Tj=Device junction temperature (calculate using Ta + Rthja*P)
     (C)
```

Microcircuits GaAs

```
function  Hp=microcircuits_GaAs(PE_type,PQ_type,pkg,y,Np,DT,C,Tj),
-----------------------------------------------------------------------
     Hp=microcircuits_GaAs(PE_type,PQ_type,pkg,y,Np,DT,C,Tj)
     PE_type: 1  2  3  4  5  6   7   8   9   10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF AUC AUF ARW  SF  MF  ML  CL
     Quality class:
    PQ_type= 1 For class S categories:a.procured in full accordance
               with MIL-M-38510.class S requirements
             2 For class B categories: a.procured in full accordance
               with MIL-M-38510.class B requirements
             3 For class B-1 category:-procured to a in production
               MIL drawing. DESC drawing or the government approved
               documentation.
             4 Other commercial or unknown screening levels
```

```
     pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                         gnd array(PGA)1.SMT(leaded and nonleaded)
                       2 For DIPs with glass
                       3 For fatpacks with axial leads on 50 Mil
                         centers
                       4 For cans
                       5 For nonhermetic DIPs,PGA,SMT(Leaded and
                         nonleaded)
     y=Number of years the device has been in production
     Np=Number of functional pins
     DT : Device type :-  MMIC devices:-
    1 For low noise & low power(<=100mW)
    2 For driver & high power(>100mW)
    3 For unknown
     Digital devices:-
    4 For Digital devices
      C:Complexity(No. of elements)
      Tj=Device junction temperature (calculate using Ta + Rthja*P)
      (C)
```

Microcircuits Linear

```
function Hp=microcircuits_linear_ic(PE_type,PQ_type,pkg,y,Np,N,Tj),
----------------------------------------------------------------------
     Hp=microcircuits_linear_ic(PE_type,PQ_type,pkg,y,Np,N,Tj)
     PE_type: 1  2  3  4  5  6   7   8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF AUC  AUF  ARW  SF  MF  ML  CL
     Quality class:
    PQ_type= 1 For class S categories:a.procured in full accordance
               with MIL-M-38510.class S requirements
             2 For class B categories: a.procured in full accordance
               with MIL-M-38510.class B requirements
             3 For class B-1 category:-procured to a in production
               MIL drawing. DESC drawing or the government approved
               documentation.
             4 Other commercial or unknown screening levels
     pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                         gnd array(PGA)1.SMT(leaded and nonleaded)
                       2 For DIPs with glass
                       3 For fatpacks with axial leads on 50 Mil
                         centers
                       4 For cans
                       5 For nonhermetic DIPs,PGA,SMT(Leaded and
                         nonleaded)
```

```
     y=Number of years the device has been in production
     Np=Number of functional pins
     N=Number of transistors in the IC
     Tj=Device junction temperature (calculate using Ta +
        Rthja*P)(C)
```

Microcircuits Logic Array

```
function Hp=microcircuits_logic_array(PE_type,PQ_type,pkg,y,Np,
family,N,Tj),
----------------------------------------------------------------------
     Hp=microcircuts_logic_array(PE_type,PQ_type,pkg,y,Np,family,
        N, Tj)
     PE_type: 1  2  3  4  5  6   7   8   9   10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF AUC AUF ARW  SF  MF  ML  CL
     Quality class:
    PQ_type= 1 For class S categories:a.procured in full accordance
               with MIL-M-38510.class S requirements
             2 For class B categories: a.procured in full accordance
               with MIL-M-38510.class B requirements
             3 For class B-1 category:-procured to a in production
               MIL drawing. DESC drawing or the government approved
               documentation.
             4 Other commercial or unknown screening levels
     pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                         gnd array(PGA)1.SMT(leaded and nonleaded)
                       2 For DIPs with glass
                       3 For fatpacks with axial leads on 50 Mil
                         centers
                       4 For cans
                       5 For nonhermetic DIPs,PGA,SMT(Leaded and
                         nonleaded)
     y=Number of years the device has been in production
     Np=Number of functional pins
     family : Bipolar PLA/PAL
                  1 For TTL,STTL,CML,HTTL,FTTL,DTL,FCL,ASTTL
                  2 For BiCMOS,LSTTL,LTTL,ALSTTL
                  3 For III,ISL, MOS, PLA/PAL
                  4 MOS, CMOS
     N=Number of gates in the IC
     Tj=Device junction temperature (calculate using Ta +
        Rthja*P)(C)
```

Microcircuits Memories

```
function Hp=microcircuits_memories(PE_type,PQ_type,pkg,y,Np,family,
B,MT,C,ECC,Tj),
-----------------------------------------------------------------------
 Hp=microcircuits_memories(PE_type,PQ_type,pkg,y,Np,family,B,MT,C,
    ECC,Tj)
     PE_type: 1  2  3  4  5  6   7   8   9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF AUC AUF  ARW  SF  MF  ML  CL
     Quality class:
    PQ_type= 1 For class S categories:a.procured in full accordance
               with MIL-M-38510.class S requirements
             2 For class B categories: a.procured in full accordance
               with MIL-M-38510.class B requirements
             3 For class B-1 category:-procured to a in production
               MIL drawing. DESC drawing or the government approved
               documentation.
             4 Other commercial or unknown screening levels
     pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                         gnd array(PGA)1.SMT(leaded and nonleaded)
                       2 For DIPs with glass
                       3 For fatpacks with axial leads on 50 Mil
                         centers
                       4 For cans
                       5 For nonhermetic DIPs,PGA,SMT(Leaded and
                         nonleaded)
     y=Number of years the device has been in production
     Np=Number of functional pins
     family :       MOS:-
                          1 For ROM
                          2 For PROM,UVEROM,EEPROM & EAPROM
                          3 For DRAM
                          4 For SRAM(MOS & BiCMOS)
                BIPOLAR:-
                          5 for ROM & PROM
                          6 for SRAM
     B=Memory size,B(bits)= 1 For up to 16K
                            2 For 16K<B<=64K
                            3 For 64K<B<=256K
                            4 For 256K<B<=1M
     MT=Memory device type:-1 For fotox
                            2 For textured poly
```

```
                          3 For other devices(except fotox and
                            textured poly EEPROMS)
     C=Total no of programming cycles over EEPROM life.C= 1 to
       500000.
     ECC:Error correction code options:-1 For no on-chip ECC
                                       2 For on-chip hamming code
                                       3 For two-needs-one
                                         redundant cell approach
     Tj=Device junction temperature (calculate using Ta +
        Rthja*P)(C)
```

Microcircuits uPs

```
function Hp=microcircuits_Microprocessor(PE_type,PQ_type,pkg,y,Np,
family,N,Tj),
-----------------------------------------------------------------------
     Hp=microcircuits_Microprocessor(PE_type,PQ_type,pkg,y,Np,
     family,N,Tj)
     PE_type: 1  2  3  4   5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS  NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
     Quality class:
    PQ_type= 1 For class S categories:a.procured in full accordance
               with MIL-M-38510.class S requirements
             2 For class B categories: a.procured in full accordance
               with MIL-M-38510.class B requirements
             3 For class B-1 category:-procured to a in production
               MIL drawing. DESC drawing or the government approved
               documentation.
             4 Other commercial or unknown screening levels
     pkg :Package type:1 For hermetic:dips w/solder or weld seal,pin
                         gnd array(PGA)1.SMT(leaded and nonleaded)
                       2 For DIPs with glass
                       3 For fatpacks with axial leads on 50 Mil
                         centers
                       4 For cans
                       5 For nonhermetic DIPs,PGA,SMT(Leaded and
                         nonleaded)
     y=Number of years the device has been in production
     Np=Number of functional pins
     family : 1 For TTL,STTL,CML,HTTL,FTTL,DTL,FCL,ASTTL
              2 For BiCMOS,LSTTL,LTTL,ALSTTL
              3 For III,ISL
              4 For digital MOS,CMOS
```

```
     N=Number of Bits(up to 32 bits)
     Tj=Device junction temperature (calculate using Ta + Rthja*P)
     (C)
```

Microcircuits SAW

```
function Hp=microcircuits_SAW(PE_type,PQ_type),
----------------------------------------------------------------------
     Hp=microcircuits_SAW(PE_type,PQ_type)
     PE_type: 1  2  3  4  5  6   7   8   9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF AUC AUF  ARW  SF  MF  ML  CL
     Quality class:screening level:-
    PQ_type=1 For 10 temp cycles(-55C to 125C)with end point
              electrical tests at temp extremes
            2 For none beyond best commercial practices
```

Microcircuits VLSI

```
function Hp=microcircuits_VLSI(PE_type,PQ_type,NP,pkg,CFT,Tj,pt,mpt,
VTH,XS,DA1),
----------------------------------------------------------------------
     VHSIC/VHSIC like and VLSI CMOS
     Hp=microcircuits_VLSI(PE_type,PQ_type,NP,pkg,CFT,Tj,pt,mpt,VTH,
        XS,DA1)
     PE_type: 1  2  3  4  5   6   7   8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU  AIC AIF AUC  AUF  ARW  SF  MF  ML  CL
     Quality class:
    PQ_type= 1 For class S categories:a.procured in full accordance
               with MIL-M-38510.class S requirements.
             2 For class B categories: a.procured in full accordance
               with MIL-M-38510.class B requirements.
             3 For class B-1 category:-procured to a in production
               MIL drawing. DESC drawing or dher government approved
               documentation.
             4 Other commercial or unknown screening levels
     NP=Number of package pins
     pkg:-Package type:  1 For DIP
                         2 For pin grid array
                         3 For chip carrier(surface mount
                           technology)
     CFT:-Package type correction factor type:1 For hermetic
                                              2 For nonhermetic
     Tj=Device junction temperature (calculate using Ta +
        Rthja*P)(C)
```

```
     pt:-Part type:1 For logic and custom
                   2 For gate array and memory
    mpt:-Manufacturing process type:
     1 For QML(qualifed manufacturer list) or QPL(qualifed product
       list)
     2 For non QML or non QPL
     VTH=ESD susceptibility(volts)
     XS=Feature size(microns)(.80 to 1.25)
     DA1:Total Scribed Chip Die Area in cm2
       (die area conversion cm^2=MIL^2+155000)
```

Opto-electronics OPT

```
function Hp=optoelectronics(type,PE_type,PQ_type,Tj,Hb_type,a1,
a2,a3,a4,a5),
----------------------------------------------------------------------
     Type 1 : Detectors,isolators & emitters
     Hp=optoelectonics(type,PE_type,PQ_type,Tj,Hb_type)
     PE_type: 1  2  3  4  5   6    7    8    9   10   11  12  13  14
Env. factor : GB GF GM NS NU  AIC  AIF  AUC  AUF ARW  SF  MF  ML  CL
                  PQ_type:   1      2    3    4      5
     PQ_type=Quality factor; JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
     Hb_type:-Base failure rate for optoelectonics
     Hb_type                        Optoelectonics types
              Photodetectors
       1                            Photo transistor
       2                            Photo diode
              Opto-isolators
       3                            Photo diode output,single device
       4                            Phototransistor output ,single
                                    device
       5                            Photodarlington output,single
                                    device
       6                            Light sensitive resistor,single
                                    device
       7                            Photodiode output,dual device
       8                            Phototransistor output,dual
                                    device
       9                            Photodarlington output,dual
                                    device
      10                            Light sensitive resistor,dual
                                    device Emitters
```

```
       11                              Infrared light emitting
                                       diode(IRLD)
       12                              Light emitting diode(LED)
-------------------------------------------------------------------
     Type 2 : Alphanumeric display
     Hp=optoelectonics(type,PE_type,PQ_type,Tj,Hb_type,C,Hic)
     PE_type: 1  2  3  4   5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF GM NS  NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
                    PQ_type:    1     2     3     4      5
     PQ_type=Quality factor; JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
     Hb_type= 1. For segment displays
              2. For diode array displays
     C=Number of characters= 1 to 15
     Hic=1 For displays with logic chip
        =2 For displays without logic chip
-------------------------------------------------------------------
     Type 3 : Laser diode
     Hp=optoelectonics(type,PE_type,PQ_type,Tj,Hb_type,PA_type,I,
     Pr,Ps,DC)
     PE_type: 1  2  3  4   5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF GM NS  NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
     PQ_type=Quality factor; 1=Hermetic package
                             2=Nonhemertic with facet coating
                             3=Nonhemertic without facet coating
     Tj=Junction temperature(C)
     Hb=Base failure rate; Hb_type = 1 For GaAs/AI GaAs
                                     2 For In GaAs/In GaAsP
     PA_type=Application factor :-1 for CW & 2 for pulsed.
     I=Forward peak current(Amps),<=25
     Pr=Required optical power output(mW)  0<Pr/Ps<=0.95
     Ps=Rated opticalpower output(mW)
     DC=Duty cycle 0.1 to 1
```

Potentiometer POT

```
function Hp=potentiometer(type,PE_type,PQ_type,a1,a2,a3,a4,a5,a6,a7,
a8),
-------------------------------------------------------------------
     Type 1 : Thermistor (RTH)
     Hp=potentiometer(type,PE_type,PQ_type,Hb_type,R)
```

```
     PE_type: 1  2   3  4   5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF  GM NS  NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
                    PQ_type:    1         2
     PQ_type=Quality factor: MIL-SPEC   Lower
     Hb=Base failure rate: Hb_type:- 1 for Bead, 2 for Disk & 3 for
     Rod.
     R=Resistance in ohm
-----------------------------------------------------------------------
     Type : 2 Variable, Wirewound (RTR, RT)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR)
     PE_type: 1  2  3  4   5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF GM NS  NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
                    PQ_type:   1   2   3   4       5           6
     PQ_type=Quality factor:   S   R   P   M   MIL-R-27208   Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
     Pa = Power Dissipation
     VR(rated)=40 Volts for RT 26 and 27
              =90 Volts for RTR 12, 22 and 24; RT 12 and 22
-----------------------------------------------------------------------
     Type 3 : Variable, Wirewound, Precision (RR)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR,
     PC_type)
     PE_type: 1  2   3  4   5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF  GM NS  NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
                    PQ_type:    1         2
     PQ_type=Quality factor: MIL-SPEC   Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
     Pa = Power Dissipation
```

```
     VR(rated)=250 Votts for RR0900, RRI 100,RR1300, RR2000,
               RR3000,RR31OO, RR3200, RR3300,RR3400, RR3500|
     VR(rated)= 423 Volts for RR3600, RR3700
     PC=Construction class factor:  construction class type:
      PC_type=1 RR0900A2A9J103
             2   3
             3   4
             4   5
-----------------------------------------------------------------------
     Type 4 : Variable, Wirewound, Semiprecision (RA, RK)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR)
     PE_type:  1 2  3  4  5  6    7    8   9   10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC AUF ARW  SF  MF  ML  CL
                    PQ_type:    1         2
     PQ_type=Quality factor: MIL-SPEC   Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
     Pa = Power Dissipation
     VR(rated)=50 Votts for RA1o
              =75 Volts for RA20X-XC, F
              =130 Volts for RA30X-XC, F
              =175 Volts for RA20X-XA
              =275 Volts for RK09
              =320 Volts for RA30X-XA
-----------------------------------------------------------------------
     Type 5 : Variable, Wirewound, Power (RP)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR,
     PC_type)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
                    PQ_type:    1       2
     PQ_type=Quality factor: MIL-SPEC   Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
```

```
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
     Pa(applied) = Power Dissipation
     VR(rated)=250 Volts for RP06,RP10
              =500 Volts for others
     PC=Construction class factor:
     construction class type:PC_type=1 Enclosed
                                     2 Unenclosed
---------------------------------    Type 6 : Variable, Nonwirewound
                                                (RJ, RJR)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR)
     PE_type: 1  2  3  4   5   6   7   8   9    10   11  12  13  14
Env. factor : GB GF GM NS  NU  AIC AIF AUC AUF  ARW  SF  MF  ML  CL
                    PQ_type:   1   2   3    4       5         6
     PQ_type=Quality factor:   S   R   P    M   MIL-R-22097 Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
     Pa(applied) = Power Dissipation
     VR(rated)=200 Volts for RJ and RJR26;RJ and RJR50
              =300 Volts for All Others
-----------------------------------------------------------------------
     Type 7 : Variable, Composition (RV)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR)
     PE_type: 1  2   3  4   5   6   7   8   9   10   11  12  13  14
Env. factor : GB GF  GM NS  NU  AIC AIF AUC AUF ARW  SF  MF  ML  CL
                    PQ_type:    1         2
     PQ_type=Quality factor: MIL-SPEC   Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
```

```
     Pa(applied) = Power Dissipation
     VR(rated)= 500 Volts for RV4X-XA & XB
              = 500 Volts for 2RV7X-XA & XB
              = 350 Volts for RV2X-XA & XB
              = 350 Volts for RV4X-XA & XB
              = 350 Volts for RV5X-XA & XB
              = 350 Volts for RV6X-XA & XB
              = 250 Volts for RV1 X-XA & XB
              = 200 Volts for All Other Types
-----------------------------------------------------------------------
     Type 8 : Variable, Nonwirewound, Film & Precision (RQ, RVC)
     Hp=potentiometer(type,PE_type,PQ_type,R,S,T,Ntaps,R1,Pa,VR,
     Hb_type)
     PE_type: 1  2  3  4  5   6    7    8    9    10  11  12  13  14
Env. factor : GB GF GM NS NU  AIC  AIF  AUC  AUF  ARW SF  MF  ML  CL
                    PQ_type:    1         2
     PQ_type=Quality factor: MIL-SPEC   Lower
     R=Resistance in ohm
     S=Ratio of operating power to rated power
     T=Ambient temperature(C)
     Ntaps=Number of potentiometer taps including the wiper and
           terminations
     R1= Nominal Total Potentiometer Resistance
     Pa(applied) = Power Dissipation
     VR(rated)= 250 Votts for RQ090, 110, 150, 200,300
              = 500 Vofts for RQ100,160,210
              = 350 Volts for RVC5,6
     Hb=Base failure rate:
      Hb_type:1 For MIL-R-39023 (RQ)Variable, Nonwirewound, Film,
                Precision
              2 For MIL-R-23285 (RVC) Variable, Nonwirewound, Film
```

Quartzcrystals QZ

```
function Hp=Quartzcrystals(PE_type,PQ_type,F),
-----------------------------------------------------------------------
     Hp=Quartzcrystals(PE_type,PQ_type,F)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
     PQ=Quality factor:PQ_type : 1 For MIL-SPEC
                               : 2 For lower
     F=Frequency(MHZ)
```

Resistor R

```
function Hp=resistor(type,PE_type,PQ_type,T,S,Pd),
SYNTAX:
 Hp=resistor(type,PE_type,PQ_type,T,S,Pd)
```

type	Resistor style	Spec MIL-R-	Description
1.	RC	11	Fixed, composition(insulated)
2.	RCR	39008	Fixed, composition(insulated) Est.R
3.	RL	22684	Fixed, flm, insulated
4.	RLR	39017	Fixed, flm, insulated, Est. R
5.	RN(R,C or N)	55182	Fixed, flm, Est. R
6.	RM	55342	Fixed, flm, Chip, Est. R
7.	RN	10509	Fixed, flm (high stability)
8.	RD	11804	Fixed, flm (Power type)
9.	RZ	83401	Fixed, flm, resistor networks
10.	RB	93	Fixed, wirewound (accurate)
11.	RBR	39005	Fixed, wirewound (accurate) Est.R
12.	RW	26	Fixed, wirewound (power)
13.	RWR	39005	Fixed, wirewound(power) Est. R
14.	RE	18546	Fixed, wirewound(power, chassis mnt)
15.	RER	39009	Fixed, wirewound(power, chassis mnt) Est. R
16.	RTH	23648	Thermistor, insulated
17.	RT	27208	Variable, wirewound (Lead screw activated)
18.	RTR	39015	Variable, wirewound (Lead screw activated) Est.R
19.	RR	12934	Variable, wirewound, precision
20.	RA	19	Variable, wirewound (low operating temp.)
21.	RK	39002	Variable, wirewound, semi-precision
22.	RP	22	Wirewound, power type
23.	RJ	22097	Variable, non-wirewound
24.	RJR	39035	Variable, non-wirewound Est. R
25.	RV	94	Variable, composition

```
  26.      RQ            39023         Variable, non-wirewound
                                       precision
  27.      RVC           23285         Variable, non-wirewound
-----------------------------------------------------------------------
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
            PQ_type:   1     2    3    4     5       6
     Quality factor:   S     R    P    M  MIL-R-xx  Lower
     T=Resistor case temperature or ambient component temperature(C)
     S=Ratio of operating power to rated power
     Pd=Operating power dissipation in resistor
```

Relays RY

```
function Hp=relays(type,PE_type,PQ_type,Hb_type,TA,S,PC_type,
PL_type,CYRenv,CYR,CT),

-----------------------------------------------------------------------
     Type 1 : mechanical
 Hp=relays(type,PE_type,PQ_type,Hb_type,TA,S,PC_type,PL_type,
 CYRenv,CYR,CT)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
                PQ_type:  1  2  3   4   5   6       7
                          8
 PQ_type=Quality factor:  R  P  X   U   M   L   Non-est.rel
                          Commercial
     Hb=Base failure rate: Hb_type:1 for T=85 C max rated
                                  :2 for T=125 C max rated
     TA=Ambient temperature (C)
     S=Operating load current/rated resistive load current
     PC_type: Contact form factor  :1 for SPST
                                   :2 for DPST
                                   :3 for SPDT
                                   :4 for 3PST
                                   :5 for 4PST
                                   :6 for DPDT
                                   :7 for 3PDT
                                   :8 for 4PDT
                                   :9 for 6PDT
     PL_type :Load stress factor for    :1 for resistive
                                        :2 for inductive
                                        :3 for lamp
```

CYRenv= quality type for cycle rate :1 for MIL-SPEC
:2 for lower quality

CYR=Cycle rate (cycles per hour)

CT: construction type – choose from the list given below

Contact rating	Construction type Application type	CT
Signal current (low mV and mA)	Dry circuit	1 Armature(long) 2 Dry reed 3 Mercury wetted 4 Magnetic latching 5 Balanced armature 6 Solenoid
0 - 5 Amp	General Purpose	7 Armature(long) 8 Balanced armature 9 Solenoid
	Sensitive (0 - 100 mV)	10 Armature(long & short) 11 Mercury wetted 12 Magnetic latching 13 Meter movement 14 Balanced armature
	Polarized	15 Armature(short) 16 Meter movement
	Vibrating reed	17 Dry reed 18 Mercury wetted
	High speed	19 Armature(balanced &short 20 Dry reed
	Thermal time delay	21 Bimetal
	Electronic time delay(non thermal)	22
	Latching magnetic	23 Dry reed 24 Mercury wetted 25 Balanced arma-ture

		26 Vaccum (glass)
	High voltage	27 Vaccum (ceramic)
5 - 20 Amp		28 Armature(long & short)
		29 Mercury wetted
	Medium power	30 Magnetic latching
		31 Mechanical latching
		32 Balanced armature
		33 Solenoid
		34 Armature(short)
		35 Mechanical latching
25 - 600 Amp	Contactors (high current)	36 Balanced armature
		37 Solenoid

Type 2 : solid state and time delay

Hp=relays(type,PE_type,PQ_type,Hb_type)

PE_type: 1 2 3 4 5 6 7 8 9 10 11 12 13 14

Env. factor : GB GF GM NS NU AIC AIF AUC AUF ARW SF MF ML CL

PQ_type: 1 2

PQ_type=Quality factor: MIL-SPEC Lower

Hb=Base failure rate:Hb_type:-1 For solid state
2 For solid state time delay
3 For hybrid

Rotating Device ROT

function Hp=rotatingdevice(type,a1,a2,a3,a4,a5),

Type 1 : Motors

Hp=rotatingdevice(type,T_type,M_type,T,LC,H)

T_type : 1. ambient temperature(C)
2. cycled temperature

M_type : 1. Electrical (general)
2. Sensor
3. Servo
4. Stepper

```
          T : for T_type=1 ->ambient temperature (deg.C)
              for T_type=2 ->[temp1, temp2, temp3,....]
          LC : system design life cycle in hours or
               the motor preventive maintenance interval in hours
          H : for T_type = 2 ONLY
              [time @ temp1, time @ temp2, time @ temp3, ....]
-----------------------------------------------------------------------
     Type 2 : Synchros and resolvers
     Hp=rotatingdevice(type,PE_type,PN,DT,SIZE,TF)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
     PN=Number of brushes 2, 3,  4
     DT=Device type: 1 For synchro
                   : 2 For resolver
     SIZE:1 For size 8 or smaller
         :2 For size 10-16
         :3 For size 18 or larger
     TF=Frame temperature(C)
         (if TF is unknown assume TF=40 C+ambient temperature)
-----------------------------------------------------------------------
     Type 3 : Elapsed time meters
     Hp=rotatingdevice(type,PE_type,PT_type,Hb_type)
     PE_type: 1  2  3  4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
     PT=Temperature stress factor:-  Operating T(C)/rated T(C)
        PT_type
                                    0 to 0.5                  1
                                    0.6                       2
                                    0.8                       3
                                    1.0                       4
     Hb=Base failure rate:  type                 Hb_type
                            A.C.                   1
                            Inverter driven        2
                            Commutator D.C.        3
```

Switches SW

```
function Hp=switches(type,PE_type,PQ_type,PC_type,PL_type,S),
-----------------------------------------------------------------
  type: switch type
             1. Centrifugal
             2. Dual in line package
```

```
             3. Limit
             4. Liquid level
             5. Microwave (waveguide)
             6. Pressure
             7. Pushbutton
             8. Reed
             9. Rocker
            10. Rotary
            11. Sensitive
            12. Thermal
            13. Thumbwheel
            14. Toggle
            15. Circuit breaker - (not used as power on/off switch)
                (magnetic, thermal and thermal-magnetic switches)
            16. Circuit breaker - (also used as power on/off switch)
                (magnetic, thermal and thermal-magnetic switches)
-----------------------------------------------------------------------
     PE_type: 1  2  3  4  5  6    7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                  PQ_type:       1          2
    PQ_type=Quality factor:    MIL-SPEC    Lower
    PC_type :Contact confguration factor  :1 for SPST
                                          :2 for DPST
                                          :3 for SPDT
                                          :4 for 3PST
                                          :5 for 4PST
                                          :6 for DPDT
                                          :7 for 3PDT
                                          :8 for 4PDT
                                          :9 for 6PDT
    PL_type :Load stress factor         :1 For resistive load
                                        :2 For inductive load
                                        :3 For lamp load
    S=Operating load current/rated resistive load current
```

Transformer XFM

```
function Hp=transformer(type,PE_type,PQ_type,THS)
-----------------------------------------------------------------------
     Transformers
     Hp=transformer(type,PE_type,PQ_type,THS)
     type:     1. Flyback (<20 volts)
               2. Audio (15-20KHz)
```

```
              3. Low power pulse (Peak pwr < 300W and avg. pwr <5W)
              4. High power, high power pulse (peak power >=300W
                 and avg.pwr >=5W)
              5. RF(10K - 10M Hz)
     PE_type: 1   2   3   4   5   6   7    8   9   10  11 12  13  14
Env. factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                  PQ_type:     1                 2
     PQ_type=Quality factor:  MIL-SPEC           Lower
     THS=Hot spot temperature(C)
```

Transistor FET-M

```
function Hp=transistor(type,PE_type,PQ_type,Tj,a1,a2,a3,a4,a5,a6,a7,
a8),
------------------------------------------------------------------------
     Type 1 : Transistors,low frequency,bipolar
     Hp=transistors(type,PE_type,PQ_type,Tj,Pr,VCE,VCEO,PA_type)
     PE_type: 1  2  3  4  5   6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                  PQ_type:    1     2    3    4      5
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
     Pr=Rated power in watt
     VCE= Collector to emitter voltage
     VCEO=Collector to emitter voltage,base open
     PA=Application factor=>PA_type=>1 for linear amplifcation
        2 for switching
------------------------------------------------------------------------
     Type : 2 Transistors,low frequency,SI FET
     Hp=transistors(type,PE_type,PQ_type,Tj,Pr,Hb_type)
     PE_type: 1  2  3  4  5  6    7     8   9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                  PQ_type:    1     2    3    4      5
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
     Pr=Rated power in watt
     Hb_type=1 For MOSFET
             2 For JFET
```

```
-----------------------------------------------------------------------
     Type 3 : Transistors,unijunction
     Hp=transistors(type,PE_type,PQ_type,Tj)
     PE_type: 1  2  3  4  5  6    7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                    PQ_type:    1     2    3    4      5
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
-----------------------------------------------------------------------
     Type 4 : Transistors,low noise,high frequency,bipolar
     Hp=transistors(type,PE_type,PQ_type,Tj,Pr,VCE,VCEO)
     PE_type: 1  2  3   4  5  6   7    8    9    10   11  12  13  14
Env. factor : GB GF GM  NS NU AIC AIF  AUC  AUF  ARW  SF  MF  ML  CL
                    PQ_type:    1     2    3    4
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER
     Tj=Junction temperature(C)
     Pr=Rated power in watt
     VCE= Collector to emitter voltage
     VCEO=Collector to emitter voltage,base open
-----------------------------------------------------------------------
     Type 5 : Transistors,high power,high frequency,bipolar
Hp=transistors(type,PE_type,PQ_type,Tj,PA_type,VCE,P,F,PT_type,
   BVCES,… PM_type,DF)
     PE_type: 1  2  3  4  5  6    7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                    PQ_type:    1     2    3    4
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER
     Tj=Junction temperature(C)
     PA_type =1 For CW application (DF=duty factor not required)
              2 For pulsed application (DF=duty factor required)
     VCE= Collector to emitter voltage
     P=Output power(W)
     F=Frequency(GHz)
     PT_type=1 For gold metallization
             2 For aluminum metallization
     BVCES=Collector-emitter breakdown voltage with base shorted to
           emitter(volts)
                        PM_type=>          1             2        3
     PM=Matching network factor=>  Input and output   Input     None
```

```
     Only if PA_type=2 enter DF.
     DF=Duty factor=[1 5 10 15 20 25 30]
-----------------------------------------------------------------------
    Type 6 : Transistors,high frequency,GaAs FET
    Hp=transistors(type,PE_type,PQ_type,Tj,PA_type,P,F,PM_type)
     PE_type: 1  2  3  4  5  6    7    8    9    10   11  12  13  14
Env. factor : GB GF GM NS NU AIC  AIF  AUC  AUF  ARW  SF  MF  ML  CL
                  PQ_type:   1     2    3    4
    PQ_type=Quality factor: JANTXV JANTX JAN LOWER
    Tj=Junction temperature(C)
    PA_type=Application factor: 1 For all low power and pulsed=1
                                2 For  CW=4
    P=Average output power(Watts)
    F=Frequency(GHz)
    PM_type=Matching network factor=>1 For input and output
                                    2 For input
                                    3 For none
-----------------------------------------------------------------------
     Type 7 : Transistors,high frequency,SI FET
     Hp=transistors(type,PE_type,PQ_type,Tj,Hb_type)
     PE_type:  1  2  3  4  5  6   7   8   9    10   11   12  13  14
Env. factor : GB GF GM  NS NU AIC AIF AUC AUF  ARW  SF   MF  ML  CL
                  PQ_type:    1     2    3    4
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER
     Tj=Junction temperature(C)
                Hb_type=     1            2
     Hb=Base failure rate=   MOSFET       JFET
```

Tubes TUB

```
function Hp=tubes(type,PE_type,a1,a2,a3,a4,a5),
-----------------------------------------------------------------------
                     SELECT THE TUBE TYPE
           1. Tubes,all types except twt and magnetron
           2. Tubes,traveling wave
           3. Tubes,magnetron
-----------------------------------------------------------------------
-----------------------------------------------------------------------
     Type 1: Tubes, all types except twt and magnetron
     Hb=Base failure rate:Hb_type:1 For pulsed Klystrons
                                  2 For any other pulsed Klystrons
```

```
                                  3 For CW
                                  4 For any other CW
-------------------------------------------------------------------
-------------------------------------------------------------------
    Type 1:- Hb_type 1 : Pulsed Klystrons
    Hp=tubes(type,PE_type,T,Hb_type,PK_type)
                PE_type:  1   2   3   4   5   6    7    8    9    10
                          11  12  13  14  |
PE=Environment factor :   GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW
                          SF  MF  ML  CL  |
    T=Number of years since introduction to feld use
    Hb_type=1
    ---------------------------------------------------------------
    | PK_type |                              |   Tube type
    |---------|------------------------------|----------------------
    |   1     |                              |Triode,tetrode,pentode
    |   2     |        Reciever              |Power rectifer
    |   3     |                              |CRT
    |   4     |                              |Thyratron
    |---------|------------------------------|----------------------
    |   5     |     Crossed feld amplifer    |QK681
    |   6     |                              |SFD261
    |---------|------------------------------|----------------------
    |   7     |                              |2041
    |   8     |      Pulsed gridder          |6952
    |   9     |                              |7835
    |---------|------------------------------|----------------------
    |   10    |                              |Triode,peak
                                               pwr<=200KW,Avg
    |         |                              |Pwr<=2KW,freq<=200MHz
    |   11    |        Transmitting          |Tetrode & pentode,peak
                                               pwr
    |         |                              |<=200KW,Avg pwr<=2KW,
                                               Freq<=200MHz|    |
    |   12    |                              |If any of the above
                                               limit exceeded |    |
    ----------------------------------------|----------------------
    |   13    |                              | Antimony
                                               trisulfde(Sb2S3)
    |         |       Vidicon                | Photo conductive
                                               material
    |   14    |                              | Silicon diode array
    |         |                              | Photo conductive
                                               material
```

15		VA144
16		VA145E
17	Twystron	VA145H
18		VA913A
19		4KMP10000LF
20		8568
21		L3035
22		L3250
23	Klystron pulsed	-
24		SAC42A
25		VA842
26		Z5010A
27		ZM3038A

Type 1 :- Hb_type 2 : For any other pulsed Klystrons

Hp=tubes(type,PE_type,T,Hb_type,P,F)

PE_type: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |

PE=Environment factor : GB GF GM NS NU AIC AIF AUC AUF ARW SF MF ML CL |

T=Number of years since introduction to feld use

Hb_type=2

P=Peak output power in MW 0.01<=P<=25 and p<=490F^-2.95

F=Operating frequency in GHz 0.2<=F<=6

Type 1:- Hb_type 3 : CW klystron

Hp=tubes(type,PE_type,T,Hb_type,PK_type)

PE_type: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 |

PE=Environment factor : GB GF GM NS NU AIC AIF AUC AUF ARW SF MF ML CL |

T=Number of years since introduction to feld use

Hb_type=3

```
    ------------------------------------------
    | PK_type |    Tube type                  |
    |---------|-------------------------------|
    |    1    |  Klystron low power           |
    |------------------------------------------
    |         | Klystron continuous wave      |
    |    2    |  3K3000LQ
    |    3    |  3K50000LF
    |    4    |  3K210000LQ
    |    5    |  3KM300LA
    |    6    |  3KM3000LA
    |    7    |  3KM50000PA
    |    8    |  3KM50000PA1
    |    9    |  3KM50000PA2
    |   10    |  4K3CC
    |   11    |  4K3SK
    |   12    |  4K50000LQ
    |   13    |  4KM50LB
    |   14    |  4KM50LC
    |   15    |  4KM50SJ
    |   16    |  4KM50SK
    |   17    |  4KM3000LR
    |   18    |  4KM50000LQ
    |   19    |  4KM50000LR
    |   20    |  4KM170000LA
    |   21    |  8824
    |   22    |  8825
    |   23    |  8826
    |   24    |  VA800E
    |   25    |  VA853
    |   26    |  VA856B
    |   27    |  VA888E
    ------------------------------------------
------------------------------------------------------------------------'
------------------------------------------------------------------------
    Type 1 :- Hb_type 4: For any other CW  Klystrons
    Hp=tubes(type,PE_type,T,Hb_type,P,F)
               PE_type: 1   2   3   4   5   6    7     8    9   10
                        11  12  13  14  |
PE=Environment factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW
                        SF  MF  ML  CL  |
    T=Number of years since introduction to feld use
    Hb_type=4
```

```
    P=Avg output power in KW 0.1<=P<=100 and p<=8.0*(10^6)*(F^-1.7)
    F=Operating frequency in MHz 300<=F<=8000
----------------------------------------------------------------------,
----------------------------------------------------------------------

    Type 2 : Tubes,traveling wave
    Hp=tubes(type,PE_type,P,F)
               PE_type: 1   2   3   4   5   6    7    8    9    10
                        11  12  13  14  |
PE=Environment factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW
                        SF  MF  ML  CL  |
    P=Rated power in watts(peak,if pulsed) 0.001<=P<=40000
    F=Operating frequency in GHz 0.1<=F<=18
----------------------------------------------------------------------,
----------------------------------------------------------------------

    Type 3 : Tubes,magnetron
    Magnetron type :M_type: 1 For pulsed magnetron              |
                            2 For CW magnetron(rated power<5KW) |
----------------------------------------------------------------------,
----------------------------------------------------------------------

    Type 3 :- magnetron type :M_type: 1 for pulsed magnetron
    Hp=tubes(type,PE_type,M_type,PC_type,R,P,F)
               PE_type: 1   2   3   4   5   6    7    8    9    10
                        11  12  13  14   |
PE=Environment factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW
                        SF  MF  ML  CL   |
    M_type=1
    PC=Construction factor :PC_type 1. CW(rated power<5KW)
                                    2. Coaxial pulsed
                                    3. Conventional pulsed
    R=Radiate hours/flament hours(0.0 to 1.0)
    P=Output power in MW  0.01<=P<=5
    F=Operating frequency in GHz 0.1<=F<=100
----------------------------------------------------------------------,
----------------------------------------------------------------------

    Type 3 : Magnetron type :M_type: 2 for CW magnetron(rated
             power<5KW)
    Hp=tubes(type,PE_type,M_type,PC_type,R)
```

```
               PE_type:  1   2   3   4   5   6    7    8    9    10
                         11  12  13  14  |
PE=Environment factor :  GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW
                         SF  MF  ML  CL  |
```

```
     M_type=2
     PC=Construction factor :PC_type 1. CW(rated power<5KW)
                                     2. Coaxial pulsed
                                     3. Conventional pulsed
     R=Radiate hours/flament hours(0.0 to 1.0)
```

Thyristors SCR

```
function Hp=thyristors(PE_type,PQ_type,Tj,VBo,VBr,Ifrms),
-----------------------------------------------------------------------
     Thyristors,SCRs & triacs
     Hp=thyristors(PE_type,PQ_type,Tj,VBo,VBr,Ifrms)
     PE_type:  1   2   3   4   5   6    7    8    9    10
               11  12  13  14
Env. factor :  GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW
               SF  MF  ML  CL
                    PQ_type:    1     2    3    4      5
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
     VBo=Blocking voltage applied
     VBr=Blocking voltage rated
     Ifrms=RMS rated forward current(Amps)
```

UJT-UJT

```
function Hp=ujt(PE_type,PQ_type,Tj),
-----------------------------------------------------------------------
     Transistors,unijunction
     Hp=ujt(PE_type,PQ_type,Tj)
     PE_type: 1   2   3   4   5   6    7    8    9    10   11  12
              13  14
Env. factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF
              ML  CL
                    PQ_type:    1     2    3    4      5
     PQ_type=Quality factor: JANTXV JANTX JAN LOWER PLASTIC
     Tj=Junction temperature(C)
```

Variable Capacitor VC

```
function Hp=variablecap(type,PE_type,PQ_type,T,S,a1,a2),
----------------------------------------------------------------------
     Type 1 : Variable ceramic(CV)
     Hp=variablecap(type,PE_type,PQ_type,T,S,Hb_type)
     PE_type: 1   2   3   4   5   6    7    8    9    10   11  12
              13  14
Env. factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF
              ML  CL
                    PQ_type:       1          2
     PQ_type=Quality factor:    MIL-SPEC    Lower
     T=Ambient temperature(C)
     S=Ratio of operating power to rated power
     Hb=Base failure rate: Hb_type:1 For T=85 C max rated((MIL-C-81
                                     Styles CV 11,14,21,31,
                                     32,34,40,41)
                                   :2 For T=125 C max rated(MIL-C-81
                                     Styles CV 35, 36)
----------------------------------------------------------------------
----------------------------------------------------------------------
     Type 2 : Variable piston type(PC)
     Hp=variablecap(type,PE_type,PQ_type,T,S,Hb_type)
     PE_type:  1   2   3   4   5   6    7    8    9    10   11  12
               13  14
Env. factor :  GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF
               ML  CL
     PQ_type=Quality factor:    MIL-SPEC    Lower
     T=Ambient temperature(C)
     S=Ratio of operating power to rated power
     Hb=Base failure rate: Hb_type:1 For T=125 C max rated
                                     (MIL-C-14409 Styles G, H, J,
                                     L, T)
                                   :2 For T=150 C max rated(MIL-C-
                                     14409 Characteristic Q)
----------------------------------------------------------------------
----------------------------------------------------------------------
     Type 3 : Variable, Air trimmer
     Hp=variablecap(type,PE_type,PQ_type,T,S)
```

```
     PE_type:  1   2   3   4   5   6    7    8    9    10   11  12
               13  14
Env. factor :  GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF
               ML  CL
                    PQ_type:       1          2
     PQ_type=Quality factor:    MIL-SPEC    Lower
     T=Ambient temperature(C)
     S=Ratio of operating power to rated power
-----------------------------------------------------------------------
-----------------------------------------------------------------------
     Type 4 : Variable and fxed, gas or vaccum(CG)
     Hp=variablecap(type,PE_type,PQ_type,T,S,Hb_type,PCF_type)
     PE_type: 1   2   3   4   5   6    7    8    9    10   11  12
              13  14
Env. factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF
              ML  CL
                    PQ_type:       1          2
     PQ_type=Quality factor:    MIL-SPEC    Lower
     T=Ambient temperature(C)
     S=Ratio of operating power to rated power
     Hb=Base failure rate: Hb_type:1 For T=85 C max rated (Styles CG
                                     20,21,30,31,32,40-44, 51,60-
                                     64,67)
                                  :2 For T=100 C max rated (Styles
                                     CG 65, 66)
                                  :3 For T=125 C max rated (Style CG
                                     50)
     PCF=Confguration factor:PCF_type:-1 For fxed
                                       2 For variable
```

HYBRID Micro-HYB

```
function Hp=hybridmicro(CKT,PE_type,class,y)
-----------------------------------------------------------------------
     Hp=hybridmicro(CKT,PE_type,class,Y)
     CKT:Circuit type:- 1 For digital
                        2 For video,10MHz<1<1GHz
                        3 For microwave,f>1 GHz
                        4 For linear,f<10 MHz
                        5 For power
```

```
     PE_type: 1   2   3   4   5   6    7    8    9    10   11  12
              13  14
Env. factor : GB  GF  GM  NS  NU  AIC  AIF  AUC  AUF  ARW  SF  MF
              ML  CL
     Class:-
      1 For class S categories:-
        a.procured in full accordance with MIL-M-38510.class S
          requirements.
        b.procured in full accordance with MIL-I-38535 and appendix
          B thereto(class U).
        c.hybrids:(procured to class S requirements(quality level K)
          of MIL-H-38534).
      2 For class B categories:-
        a.procured in full accordance with MIL-M-38510.class B
          requirements.
        b.procured in full accordance with MIL-I-38535.(class Q)
        c.hybrids:procured to class B requirements(quality level
          H)of MIL-H-38534).
      3 For class B-1 category:-procured to a in production MIL
        drawing. DESC drawing or dher government approved
        documentation. |
      4 For TM1010 and TM2001 and TM5004 and TM1014 and TM2009
      5 For TM1010 or TM2001,TM5004 and TM1014 and TM2009
      6 For pre-bum in electricals TM1015 and TM5004 B level
      7 For pre-bum in electricals TM1015 and TM5004 S level
      8 For TM2020 particle impact noise detection
      9 For TM5004(or 5008 for hybrids)
     10 For TM2010/17
     11 For TM1014
     12 For TM2012
     13 For TM2009
     14 For TM5007/5013(GaAs)
     15 For TM2023
     16 Other commercial or unknown screening levels
     y:years generic device type has been in production
```

Bibliography

1. Anderson, R.T. (March 1976) *Reliability Design Handbook*, Rome Air Development Center, NY, USA.
2. Billinton, R. and Allan, R. (1983) *Reliability Evaluation and Design*, Plenum Press, NY, USA.
3. Bose, B.K. (February 1982) Adjustable Speed AC Drives – A Technology Status Review. *Proc. IEEE*, **70** (2), 116–135.
4. Colonel McLyman, T.W. (1978) *Transformer and Inductor Design Handbook*, Marcel Dekker, Inc., NY, USA.
5. Crandall, S.H., Karnopp, D.C., Kurtz, E.F. (1968) *Dynamics of Mechanical and Electromechanical Systems*, McGraw-Hill Book Company, NY, USA.
6. Demidovich, B.P. and Maron, I.A. (1976) *Computational Mathematics*, Mir Publishers, Moscow, Russia.
7. Doyle, J.C., Stein, G. (February 1981) Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis. *IEEE Trans. Automat. Control*, **AC-26**(1), 4–16.
8. Finney, D. (1980) *The Power Thyristor and its Applications*, McGraw-Hill Book Company, London, UK.
9. Franklin, G.F., Powell, J.D. and Workman, M.L. (1990) *Digital Control of Dynamic Systems*, 2nd edn, Addison-Wesley, MA, USA.
10. Friedland, B. (1987) *Control System Design – An Introduction to State Space Methods*, McGraw-Hill Book Company, Singapore.
11. Fuqua, N.B. (1987) *Reliability Engineering for Electronic Design*, Marcel Dekker, Inc., NY, USA.
12. Gottlieb, I.M. (1985) *Power Supplies – Switching Regulators, Inverters and Converters*, BPB Publications, New Delhi, India.
13. Johnson, H.W. and Graham, M. (1993) *High Speed Digital Design – A Handbook of Black Magic*, Prentice Hall, NJ, USA.
14. Kitsum, K. (1984) *Switched Mode Power Conversion – Basic Theory and Design*, Marcel Dekker, Inc., NY, USA.
15. Klion, J. (1992) *Practical Electronic Reliability Engineering*, Van Nostrand Reinhold, NY, USA.
16. Kraus, J.D. (1984) *Electromagnetics*, 3rd edn, McGraw-Hill Book Company, NY, USA.
17. Krause, P.C. (1987) *Analysis of Electric Machinery*, McGraw-Hill Book Company, NY, USA.
18. Lander, C.W. (1981) *Power Electronics*, McGraw-Hill Book Company, NY, USA.
19. Leonhard, W. (1985) *Control of Electrical Drives*, Springer Verlag, Berlin, Germany.
20. Lewis, F.L. (1992) *Applied Optimal Control and Estimation*, Prentice Hall, NJ, USA.
21. Meyer, P.L. (1970) *Introductory Probability and Statistical Applications*, Oxford and IBH Publishing Company Pvt. Ltd., New Delhi, India.
22. Military handbook: MIL-HDBK 217F (Notice 2).
23. Miller, I. and Freund, J.E. (1985) *Probability and Statistics for Engineers*, Prentice Hall of India, New Delhi, India.
24. Mohan, N., Undeland, T.M. and Robbins, W.P. (1989) *Power Electronics: Converters, Applications and Design*, John Wiley & Sons, NY, USA.
25. Mottershead, A. (1984) *Electronic Devices and Circuits – An Introduction*, Eastern Economy edn, Prentice Hall of India, New Delhi, India.

26. Murphy, J.M.D. and Turnbull, F.G. (1988) *Power Electronic Control of AC Motors*, Pergamon Press, NY, USA.
27. Nadkarni, M.A. and Bhat, S.R. (1985) *Pulse Transformer: Design and Fabrication, CEDT series*, Tata McGraw-Hill Book Company, New Delhi, India.
28. Ogata, K. (1991) *Modern Control Engineering*, Prentice Hall of India, New Delhi, India.
29. Ohno, E. (1988) *Introduction to Power Electronics*, Oxford Science Publications, Oxford, UK.
30. Pearman, R.A. (1980) *Power Electronics – Solid State Motor Control*, Reston Publishing Co. Inc., VA, USA.
31. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1986) *Numerical Recipes*, Cambridge University Press, NY, USA.
32. Ramamoorthy, M. (1991) *Thyristor and their Applications*, Affiliated East-West Press Pvt. Ltd., New Delhi, India.
33. Rashid, M.H. (1988) *Power Electronics, Circuits, Devices and Applications*, Prentice Hall, NJ, USA.
34. Shepherd, W., Hulley, L.N. and Liang, D.T.W. (1995) *Power Electronics and Motor Control*, Cambridge University Press, NY, USA.
35. Shinners, S.M. (1964) *Control System Design*, John Wiley & Sons, NY, USA.
36. Strang, G. (1986) *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, MA, USA.
37. Tarter, R.E.(1985) *Principles of Solid State Power Conversion*, Howard W. Sams and Co., Inc., IN, USA.
38. Thorborg, K. (1988) *Power Electronics*, Prentice Hall, NJ, USA.
39. Umanand, L. and Bhat, S.R. (1992) *Design of Magnetic Components for Switched Mode Power Converters*, Wiley Eastern Limited, New Delhi, India.
40. Unitrode seminar manual (1988) *Switching Regulated Power Supply Design*, Unitrode Corporation, NH, USA.
41. Vas, P. (1990) *Vector Control of AC Machines*, Clarendon Press, Oxford, UK.
42. Vithayathil, J. (1995) *Power Electronics – Principles and Applications*, McGraw-Hill Book Company, NY, USA.
43. Wildi, T. (1981), *Electrical Power Technology*, John Wiley & Sons, NY, USA.
44. William, B.W. (1992) *Power Electronics: Design, Drives, Applications*, 2nd edn, McGraw-Hill Book Company, NY, USA.

Index

A

Absorptance, 755–756
Ackerman's formula, 592, 633
AC voltage, 95
Algorithm,
  alternate algorithm, 710
  arithmetic, 707
  iterative, 709
  Kalman filter, 660
  least squares, 660, 663
  multiplication, 707–709
  square root iteration, 712
Aliasing, 521
Alpha parameter, 12
Ampere's law, 369, 370, 373
Amp-second balance, 200
Analog controller, 499
Anode, 6
  short, 44
Aluminum
  anodized, 754
  polished, 754
  rough surface, 754
Area
  cross-section, 378–379
  product, 377, 379, 391
  window, 377, 378
AT&T, 837

B

Baker's clamp, 63
Bang-bang technique, 317
BELL, 837
Belt-conveyor system, 806, 816
BJTs (bipolar junction transistors), 1
Black box, 425
Bode diagram, 552
  constraints in, 556
  controller design with, 556–561, 567
  method, 550, 551–561
  open-loop, 555
  shaping of, 557
Bode's theorem, 551, 552
Boltzmann constant, 4, 754
Boltzmann law, 4, 5
Bond, energy flux in a, 454
Bond graph, 453–473
  augmented, 462
  construction, 458–462
  drawbacks, 453
  elements of, 457
  energy, 453
  half bond, 454
  no switch element in, 467
  of an SPDT switch, 470
  power, 453, 454
Boundary limit
  circular, 727–728
  hexagonal, 727, 728–729
British Telecom, 837
Buck converter, 201, 202, 205, 450, 471–472
  inductor current ripple of a, 222
  operation of, 201
  waveforms of various signals of a, 203
  with non-idealities, 207
  with power semiconductor switches, 204
  ZCS applied to, 267–271
  ZVS applied to, 271–275
Boost converter, 206, 211, 212, 222, 472–473
  consisting of an SPDT switch, 206
  efficiency of, 612
  operation of a, 208
  output value, 206
  topology of, 206
  waveforms of various signals of a, 209
  with non-idealities, 213
Bridge
  n-level, 295
  three-level arm, 283
  three-level arm, 293
  two-arm, 296
Buck-boost converter, 213, 214, 217
  operation of, 214
  schematic of a, 219
  steady state analysis of, 217
  topology, 213
  waveforms of various signals of, 215
  with non-idealities, 219
  with power semiconductor switches, 216

C

Capacitance
  drain-gate, 79
  filter, 110
Capacitor
  current waveform, 206
  charging time constant, 69
  energy of, 69, 112
  flux walking, 248, 249, 253
  input filter, 109
  selection, 114
  size, 212
  tantalum, 114
  value of, 6, 206, 218
Capacitor current curve, 218
  negative portion of a, 218
Causality assignment, 462–463
  sequence, 462
  limitations to selection of, 457
  rules for selection of, 457, 462
Cathode, 6
CCM (continuous conduction mode), 221, 222, 472
Characteristic dimension, 743
Charge balance, 200
Chopper, 194, 290
  applications of, 196
  DC–DC converters using, 200
  dual-four quadrant, 199
  output voltage of, 196

output waveform, 192
primary side, 230
schematic of, 194–195
Chopper, dual
circuit schematic, 200
using two SPDT switches, 199
Ciruit-averaging method, 447
Circuits
anti-saturation, 63
full-wave half-controlled, 138
MOSFET drive, 70
opto-isolated, 73
S/H (sample and hold), 576
single-phase, 96
single-phase rectifier, 96
snubber, 80, 243
three-phase, 10–109
three-phase capacitor-filter rectifier, 115
three-phase rectifier, 96
transformer-isolated gate drive, 75
transistors drive, 55
voltage scaling, 149
CMOS (complementary metal oxide semiconductor), 72
hex inverting buffer, 72
logic integrated circuit, 72
non-inverting hex, 72
Conduction, 737
Conductors,
air gaps between, 403
metallic, 6
Connectors, 822–828
Controllability, 582–583
Control principles, 540–543
Control index, 348, 357
Controller
digital current, 634
optimal, 670
PI, 717–719
speed and current, 630
transfer function, 564, 565
Controller design
classical techniques, 550
guidelines for, 575–576
major issues with, 549
performance criteria, 549
robust, 679–687
state space, 550
steps, 682–685
techniques, 550–551
transfer function based, 550
Control systems
basics, 535–540
design of, 649
dynamics and augmentation, 638–641
essential tools of, 501
performance parameters of, 535
SSS performance parameters of, 535
stable, 535
unstable, 535
Convection, 736
forced, 742, 744–753
free, 742, 744–747
heat transfer by, 742–752
Converter circuits, problems with implementing, 258
Converters
AC–DC, 628
basic, 200
bidirectional power flow, 613
boost, 472–473
buck, 201, 202, 222, 450
buck-boost, 213, 214
current-controlled, 607–610
Cuck, 259–263
DC–AC, 289
DC–DC, 193
DC–DC switched mode, 193
flyback, 225, 253–257, 601
forward, 226–228
front-end, 613–619
full-bridge, 26, 248–253, 393
half-bridge, 226, 243–248
high boost, 259, 263–265
input-output relationship of, 423
isolated, 225–257
push-pull, 226, 238–243, 393
phase-controlled, 95
single-phase full-wave controlled, 95
single-phase half-wave controlled, 95
soft switching of, 265–275
step down, 201, 205
step-up, 206, 211, 216, 217
three-phase full-wave full-controlled, 95
three-phase full-wave half-controlled, 95
three-phase half-wave controlled, 95
unidirectional power flow, 613
unity power factor, 613
utility power factor, 610–613
variants, 201
ZCS, 266
ZVS, 266
Copper, polished, 754
Core
C-shaped, 404
E-shaped, 404, 405
I-shaped, 405
loss, 412
planar, 405–406
pot, 406
rod, 404
shapes, 404–406
toroidal, 406
U-shaped, 404
Cost
analysis, 770
optimization, 770
CRGO (cold rolled grain oriented) cores, 373, 379
Crosstalk, 823–825
between directly overlapping loops, 825
estimation of, 823
problems, 824–825
Cuck converter, 259–263
application of Kirchhoff's voltage law to, 261, 262

isolated, 262–263
non-isolated, 263
schematic of, 259
two operating mode, 259
Current
controlling, 607–619
drain, 79
eddy, 410
emitter, 165
holding, 39
line side, 101, 103
load, 108, 109, 201
magnetizing, 232, 400
primary, 101, 103
protection, 164
regulation, 163, 180
reverse saturation, 6
secondary winding, 101, 103
stator , 484
surge, 9
tail, 37
transformer, 64, 398–403
turn-ON, 117–120, 128
Current, stator , 484
direct axis, 481
equation, 482
quadrature axis, 481
Current regulator, analysis of, 180–184
Current, rotor
direct axis, 482
equation, 482
quadrature, 482
Current transformer, 64, 398–403
design of, 401–403
Cylindrical solid
horizontal, 746
vertical, 746–747

D

DAC (digital-to-analog conversion), 500
D'Alember's principle, 428
Darlington connection, 21
DC bus, 109, 110
DC–AC converter, 289
classification of, 289
DC–DC converters, 200
area products for, 393–394
practically realizable, 201
using chopper, 200
using inductor, 200
using capacitor, 200
DC–DC linear regulators, 149
DCM (discontinuous conduction mode), 221, 222
DC motor, 196, 465
armature-controlled, 434, 465
constant field, 435
examples of, 465–467
excited by armature supply, 466–467
first quadrant operation of a, 196
four quadrant operation of a, 197
operating zones of a, 196
second quadrant operation of a, 196
separately excited, 465–466
third quadrant operation of a, 196
DC series motor, 467
DC servomotor, 434, 435
DC steady state principles, 200
DC voltage, 95
Decoupling network, 630–631
Density
flow, 372
flux, 372–373, 407
Digital controller, 499
advantages of, 499
design of, 500
Digital integrator, 711
Digital processor, 695
Digit product, 708, 709
Diode, 2, 174
average current rating of a, 232
body, 23
classification, 8
diffusion current, 4
doping, 3
forward-biased, 2
freewheeling, 97
model, 10–11
opto-coupler, 601
opto-isolator, 602
parameters, 8
p-n junction, 3
reverse-biased, 3
selection, 113
turn-OFF of, 7
turn-ON of, 8
Zener, 36, 151, 163, 174, 180, 183, 570
Disturbance propagation, 609
Drive circuits, 55
base drive, 58
characteristics and classification of, 58
examples of, 59–70
transistor, 55
Driven inverter, 304–308
Duty cycle, 65, 77, 312
Duty ratio, 65, 66
operating, 213
Dynamic compensator, 632
Dynamics and augmentation
anti-alias filter, 636–638
controller, 638–641
plant, 636

E

Early effect, 13
Eber–Moll model, 12
Eddy current loss, 410
Electric domain, 371
Electric stress rating, 221
Electrolyte, aluminum, 114
Electromagnetic torque
equation, 490
of an induction motor, 490
Elements
capacity or C–, 454
resistance or R–, 454
inertia or L–, 454

EMI (electromagnetic interference), 111, 825–828
due to large current loop, 827
filter, 117
Emittances for common radiation surfaces, 754
Enamel paint, 754
Energy
equation, 655
flux, 454
storage, 375–376
-storage device, 369
-transfer device, 369
Environments, 842
Equation
algebraic, 463
electromagnetic torque, 489–490
state, 463–467
Equilibrium hole density, 4
Error, residual, 715
Error constant, 537
acceleration, 537
positional, 537
velocity, 537
Estimator, 591
current, 594
design, 591–598
dynamics, 591
equation of an, 591, 597
error dynamics, 592–593
full-order state, 597
gain, 591
Kalman, 670
optimal, 664
prediction, 591
reduced-order, 596–598
state, 591
Euler's backward rectangular rule, 527–528
block schematic of, 527
illustration of, 527
Euler's forward rectangular rule, 525, 526
block schematic of, 526
illustration of, 525
Exponent, 699
Exponential, 715–716

F

Failure,
mean time between, 776–778
mean time to, 770, 776–777, 795–796
Failure mode
dormant, 792
immediate, 792
of switching network, 792
Failure rate, 770, 771–773, 776
allocation, 816
basic, 837
curve, 774
general model for, 774
periods in, 774–775
variation in, 774
Failure time
probability density function, 773–776
reliability density function, 771
Faraday's law, 369, 370, 375, 378, 384, 410, 473
Fault reset, 807
Feed rate set, 807
Field-oriented control, 623, 628
flux estimators for indirect, 628
Field weakening region, 620
Filer
capacitor input, 110
common digital, 512
digital, 508–515
dual input range rectifier, 115
finite impulse response, 508, 509
first-order low-pass, 553
infinite impulse response, 508, 510
Kalman, 595
rectifier-LC, 124–128
types of, 515
First-level modeling, 166
Fixed-point system,
primary operations in, 703
problem in, 703
tracking the base point, 703
Flow, 372
density, 372
Flux, 227, 228, 232, 374
change, 384
density, 372–373, 374
estimator, 630
estimators for indirect field-oriented control, 628
in a motor, 622
magnetic, 473
rate, 386
state model, 622
swing, 411
walking, 243, 248, 249
waveform, 230
Flux density, 372–373, 374, 407
saturation, 373
Flux linkage, 484, 489
air-gap, 489, 490, 620
generalized, 490
leakage, 386
rotor, 489, 490, 622
stator, 489, 490
Flux state model
air-gap, 485, 486
generalized, 485, 486–489
rotor, 485
stator, 485, 486
Flux walking capacitor, 248, 249, 253
capacitance of, 253
rms value of the current through a, 253
Flyback converter, 253–257
development of a, 254
schematic of, 255
waveforms of various signals of, 255
Forward converter
drawbacks of, 233
dual switch, 236–238
operation of a, 233

topology, 232
waveforms, 237
with demagnetizing winding, 233–236
Forward generation, 199
mode, 198, 199
Forward motoring mode, 198
Fourier series, 124, 125, 319, 324, 356
a general waveform, 319
for line-to-line voltage waveform, 325
of a rectified waveform, 124, 125
of a waveform with quarter wave summary, 320
Four-wire connection, 162–163
Fractional part, 715
Frame
arbitrary reference, 478, 480
stator reference, 478, 479
French mathematician, 442
Frequency domain performance specifications, 681
Frequency ratio, 355
Full-bridge circuit, 249
equation, 251
implementation, 251
topology, 250
waveforms of, 252
Full-bridge configuration, 308
Full-state feedback, 584–587, 649
Function,
cosine, 714, 715
exponential, 715–716
logarithm, 716–717
sine, 714, 715

G

Gain margin, 555
Gate turn-OFF thyristor, 44
structure of a, 45
symbol of, 45
turn-OFF characteristics of a, 45
Gauge of wire, 380, 383
Glass, 754
Gummel –Poon model, 13
Generator, 799
Generation mode, 196
forward, 196
reverse, 196

H

Half-bridge configuration, 306–307
Half-bridge converter, 243–248
circuit, 245, 247
equation for, 246
operation of, 243
schematics of, 243
Harmonic index, 318, 323, 325
Harmonic elimination, 350–358
ways to implement, 354
with multiple notches, 351–352
Heat sink, 740, 741
mounting of power devices on, 760–761
size, 761
volume, 761
Heat transfer
best method for, 762–763
by conduction, 738–742
by convection, 742
by mass transport, 756
by radiation, 753–756
constants, 739
general notes on, 758–763
mechanism, 735
mixed, 758–759
modeling of, 736–738
process, 735
steady-state, 736
transient model for, 761–763
Hold
equivalence, 521, 530–533
time, 116
Hold equivalence, 530–533
zero-order, 530
Hypotenuse, 713714
Hysteresis, 373
band, 317
loop, 373, 390
loss, 373, 388, 407

I

IGBT (insulated gate bipolar transistor), 1, 30, 194, 212, 232
dynamic switching characteristics of an, 36
remarks on, 37
structure of an, 34
Inductance,
leakage, 231, 232
magnetizing, 400
total series, 117
Induction motor, 619–631, 636, 806
closed loop V/f control, 621–622
control of, 619–631
control with output feedback, 633–641
direct field-oriented control of, 625–627
example, 674–679
indirect field-oriented control of, 627–631
open loop V/f control of, 621
scalar control, 619622
speed of an, 807
steady-state model of an, 619–620
strategies for, 677–679
Inductor, 369, 375
calculation of, 205
current, 429
design of, 379
multiple-winding, 382–383
value, 205, 375, 379
Insulation factor, 404
Integer part, 715
Integrated circuits, 158
classification of, 158
example of, 158
Inverse polar plot method, 550

Inversion layer, 23
Inverter
  classification of, 290
  composition of, 290
  constraints in voltage generation by an, 319
  double-level, 291
  driven, 304–308
  features of storable core, 297–298
  multi-level, 292, 349–350
  practical, 296
  PWM (pulse-width modulated), 315
  quasi-square wave, 321, 322
  saturable base drive, 302–304
  self-driven, 297
  simple, 290
  single-level, 290
  six-step, 314–315
  three-phase, 312–318
  two-level, 629
  topologies, 290–297
  Iron, rough surface, 754
Isolated converters, 225–257
  benefits of, 225
  buck-based, 225
  buck-boost based, 225
  categories of, 225
  classification of, 226
  schematic of, 226

J
Jacobian matrix inverse, 355
Jury test, 549
Junction temperature, 735

K
Kalman estimator, 670
Kalman filter, 595, 664–668
  algorithm, 660
  implementation filter, 668
  measurement update in, 667
  recursion steps in the, 665
  steady-state, 667
  time update in, 667
Kinetic energy of a pendulum, 443
Kirchhoff's voltage law, 65, 216, 230, 261, 447, 654
  application of, 434

L
Lagrange, 442
Lagrange's method, 442–447
Lagrange's equation, 442, 444, 447
Lagrange's function, 442
Lagrangian,
  partial derivatives of the, 444, 447
  with respect to generalized coordinates, 444, 447
Laplace transformation, 431, 446, 530
Latent heat, 757
Law
  Ampere's, 369, 370, 373
  effort, 471
  Faraday's, 369, 370, 374, 375, 378, 384, 385, 410, 473
  flow, 471
Lead-lag network, 554–556
Leakage inductance, 231
  issue of, 258
  problems, 238
  solutions to, 258
Least squares
  principle, 650
  recursive , 659–668
  solution, 655–658
  weighted, 658–659
Linear differential equation, first-order, 445
Linear quadratic Gaussian, 672
Linear quadratic regulator, 543, 670–672
Linear quadratic tracker, 672–674
Linear regulators, 149, 150, 158, 162, 163
  classification of IC, 158
  efficiency of, 165
  operating principle, 150–151
  parameters of, 166
  performance quality, 165
  topologies, 151–166
Line regulation, 182
  coefficient, 172, 182
  range of, 182
Litz wire, 413
LMS algorithm, 660–664
Load
  current, 108, 109
  delta, 313
  regulation, 182
  resistance, 173
  star, 313
  three-phase, 315
  voltage, 108
Load-regulation coefficient, 172, 173
Loci
  of the roots of the characteristic equation, 575
  root, 575
Logarithm, 716–717
Lorentz force, 473

M
Magnetic amplifier, 604
Magnetic components, 369
Magnetic domain, 371
  power dissipated in, 372
Magnetic field
  energy, 413
  intensity, 388
  rotating, 474
Magnetic loss, 407–412
Magnetic retentivity, 407
Magnetics,
  general notes on, 403–415
  theory of, 300
Mantissa, 699
Mason's rule, 440–442
MATLAB, 532, 534, 549, 551, 555, 575

Mean time
between failure, 776–777
to failure, 770, 776–777, 795–796
Measurement update, 595, 665
Method
adaptive controller,
bode diagram, 550
coupled-inductor, 602
full state feedback, 550
full state feedback with estimator, 550
inverse polar plot, 550
linear quadratic, 550
linear regulator, 602
magnetic amplifier, 603
Nichols chart, 550
non-isolated converter, 603
Nyquist diagram, 550
optimal controller, 550
optimal controller with optimal estimator, 551
output feedback, 550
robust controller, 551
root locus, 550, 573–581
self-tuning, 551
state space, 582–584
Microcontroller, 695, 707
Microprocessor, 695
MIL-217, 837
Miler turn-ON, 79
MIMO (multiple-input multiple-output), 433, 542
Minimum energy principle, 653–655
mmf (magneto-motive force), 369, 371, 375
Model
average large-signal, 448, 449, 452
Eber–Moll, 12
equilibrium, 427
Gummel–Poon, 13
large-signal, 427, 447, 451
linearized small-signal, 466
small-signal, 427, 449, 452
steady state, 427, 448, 452
Modeling of systems
d-q model of, 483–490
modeling approaches, 424
physical, 424
system identification, 424
Modeling uncertainties, 681–682
Modulation
center-pulse, 327–329
control index, 318, 322, 325
end-pulse, 324–326
index, 331, 334, 357
over-, 334, 727–729
pulse-width, 318–350
quasi-square, 321
sinusoidal pulse-width, 329
space vector, 339–350
MOSFET, 193
drive circuits, 70–80
paralleling of, 29
precautions for, 29
schematic characteristics of an *n*-channel, 22
simulated model for a, 31
symbol of an n-channel, 22
symbol of a p-channel, 22
Motor, induction
assumptions for, 479
electromagnetic torque of an, 490
modeling of, 479–483
voltage equation, 479
voltage supplied to the stator of, 479
Motoring mode, 196
forward, 196, 198
reverse, 196
Multiplicand, 707, 708
Multiplier, 707
Mutual inductance, effect, 823–825

N

Nature's minimum energy principle, 413
Newton's law of motion, 442
Nichols chart method, 550
n-Material, 3
Normalized form, 699
Notch, 331
width, 356
Notch angles, 331, 351, 355
determination of, 352–355, 357
n-Region, 4
Negative rail, 156
regulator, 158
terminal, 163
Normalization and scaling factor, 704, 705
NPN transistor, 11
equivalent circuit of an, 13
operation of, 12
Number
Nusselt, 743, 744, 745
Rayleigh, 743, 744
Reynolds, 743, 744
Numerical integration method, 521, 523–530
Numeric format, 695–702
classes of a, 696
features of a, 696
fixed-point, 696
floating-point, 699–702
general notes on, 702
problems with a, 696
Nusselt number, 743, 744, 745
Nyquist diagram method, 550

O

Observability, 583–584
Observer, 591
Ohmic drop, 7
op-amp (operational amplifier), 154
Operating duty ratio, 213
Opto-coupler, 65, 600
Output feedback, 649
Overlap angle, 140–142

P

Parabola, 660
Paraboloid, 652
Part derating, 828
Pendulum
  equation of motion of an inverted, 444
  inverted, 446
  kinetic energy of a, 443
  linearized dynamic model of a, 445
  second-order differential equation, 445
  under equilibrium condition, 445
Performance index, 675
  constraint, 674
Performance parameters
  measures for, 538–540
  SSS, 539
Period
  discrete, 774
  infant mortality, 775
  useful life, 775
  wear out, 775–776
Permeability, 376, 386
  infinite, 388
Permeance, 375, 379, 382
  dependence of, 380
  infinite, 388
Perturbation, 609
pf (power factor), 55, 121–124
Phase margin, 555
Physics of failure, 853–854
  approach for, 852
  sections of, 852
PID (proportional integral derivative) controller, 562–573
  design, 567–568
  integrator windup in, 568
  parameters, 567
  practical, 570
Plant
  dynamics, 636
  input to a, 636
  output of a, 636
  state, 636
Plate
  flat, 748
  horizontal flat, 744–745
  parallel flat, 747
  rectangular, 745–746
  p-material, 3
Poles, 436, 536
Pole voltage, 312
Pole-zero map, 521, 522–523
Polynomial, 714
Positive rail, 156
  regulator, 158
Power
  active, 121
  circular, 745
  reactive, 121
Power circuits, single-phase, 129–134
Power converter, 194
Power electronics,
  applications, 1
  integral part of, 369
  system, 717
Power MOSFET, 22
  dynamic characteristics of, 25
  model for a, 30
  power dissipated in a, 30
  static characteristics of, 25
Power semiconductor switches, 1, 265, 290, 297
  circuit viewpoint of, 1
  drive viewpoint of, 2
  packaging viewpoint of, 2
  physics viewpoint of, 1
  protection viewpoint of, 1
Power switches, 55
  electrical stress rating of, 221
  high, 55
  device of, 219
  macro models of, 467
  modulated bond graph element in, 467
  selection of, 219–221
  thermal stress rating, 221
Power system
  modeling-switched, 467–473
Protection circuitry, 163
Proximity effect, 413–415, 414
  due to equal current, 414
  due to unequal current, 414, 415
Push-pull circuit, 239
Push-pull configuration, 304–306
Push-pull converters, 226, 238–243
  circuit schematic of, 239, 242
  problem with, 243
  schematic, 239
  waveforms for, 241
PWM (Pulse-width modulation), 318–350, 500, 601, 630, 720–724
  bipolar, 330, 331
  classification of, 319
  sinusoidal, 332–334
  objectives of, 318
  space vector, 724–727
  staircase, 355–357
  through phase modulation, 337
  unipolar, 330
  waveform, 721
  with harmonic elimination, 350–358
  without harmonic elimination, 319
Pynter, Prof. H.M., 453
Pythagoras theorem, 714

Q

Quadratic forms, 651–653
  practical significance of, 653
  types of, 653
Quadratic function, 652
  cases in, 652
  coefficient of, 652
  two-dimensional, 652
Quadratic performance index, 673

Quasi-square wave, 309
Fourier series representation for, 310
inverter, 309–312
output quantity of, 309, 311
regulation based on duty cycle control, 311

R

Radiation, 737
Rail
negative, 156, 157
positive, 156, 157
Rayleigh number, 743, 744
Raynolds number, 743, 744
Rectifier
circuits, 96
classification of, 95
components, 112
controlled, 95, 128–142
design of capacitor input filter, 110–116
full-wave bridge, 96, 101–103
full-wave center-tapped, 96, 99–101
half-wave, 96, 97–99
polyphase, 127
silicon-controlled, 38
un-controlled, 95, 96
Rectifier circuit, 115
three-phase capacitor-filter, 115
three-phase controlled, 134
Rectifier filter, dual input range, 115
Recursive least squares, 659
drawbacks in, 659
Redundant systems, 779, 787–799
approach in, 799
switching in, 791–795
with non-identical component, 799–804
Reflectance, 755–756
Regulation loop, 163
Regulator parameters, 172
Regulators, 585
78xx, 158
79xx, 158
efficiency of, 150
fixed, 158
IC linear, 158–160
linear, 149, 150, 154
linear quadratic, 670–672
negative, 157, 158
operating principle of, 150
positive voltage, 158
series, 149, 154–156, 157
shunt, 149, 151–154
three-pin, 158
variable, 158
Zener shunt, 151
Regulator types, 158
fixed, 158
variable, 158
Relative error, 697, 699, 700
Relative density, 697, 699
uniform, 700
Reliability
component, 805
concepts, 771–777
definition of, 771
design for, 804–813, 822–828
determination of, 771
effect of circuit design on, 805
equipment, 777
environmental effect on, 805
factors affecting, 805
inter-connection wiring in, 805
mathematical representation of, 771
modeling for, 777–788
network model, 777
prediction of, 770, 835
prediction process, 842–843
requirements, 770, 814
target specifications of, 805
Reliability prediction, 770, 835
uses of, 885
depends on, 835
Resistance,
power dissipation in, 232
total series, 117
zero, 388
Resistor,
bleeder, 127
burden, 399
power rating of, 163
Ripple output, 127
Robust controller design, 679–687
Root
loci, 575
locus method, 550, 573–581
Rotating reference frame, 619
Rotor magnetizing current, equivalent, 628
Routh–Hurwitz criterion, 549
Rule
Euler's backward rectangular, 527–528
Euler's forward rectangular, 525, 526
Tustin trapezoidal, 528–530

S

Sample time, 635
Sampling
effect of, 518–521
Nyquist, 520
Shannon, 520
time, 501
Saturable-core inverters, 298–302
behavior of, 300
designing of, 300
features of storable core, 297–298
Saturable base drive inverter, 302–304
SciLAB, 532, 534, 549, 551, 555, 575
Second-level modeling, 174, 183–184
Self-driven inverters, 297–304
Semiconductor devices, 1

Semiconductor switches, 204, 210, 216, 228
Input-output relationship of, 204
steady state analysis of, 204
with ideality constraints, 204
Sensitivity
co-, 670
complementary, 680
Series inductance effect, 825–828
Series regulator
analysis of a, 166–180
equivalent circuit of, 168
linear, 167
Shunt regulator,
modified, 154, 165
power rating of, 163
Signal
modulating, 329
switching, 329
Signal processing
analog, 499
digital, 499
Silicon-controlled rectifier, 38
biasing a, 39
circuit model of a, 42
Sine and cosine, 714–715, 719
Sinusoid, peak of a, 103, 108
SISO (single-input single-output), 433, 542, 649
Skin effect, 412–413
Skin depth, 413
Snubber,
circuits, 243
series, 83–85
shunt, 80–83
Snubber circuits, 80
turn-OFF, 80
turn-OFF dissipative, 80
turn-ON, 80
turn-ON dissipative, 80
SOAR (safe operating area), 19
forward, 19
reverse, 19
Solid cylinder, 748–749
Source
single-phase sinusoidal, 127
three-phase, 127
Space factor, 403
Space vector, 340, 473, 475
amplitude of a, 345
flux linkage, 489
general, 476
in a coplanar vector space, 343
in orthogonal coordinates, 476
in polar coordinates, 477
modeling, 473–490
PWM, 724–727
resultant, 475
resultant stator current, 475
rotation of, 344
transformation, 476–478
Space vector modulation, 339–350
features of, 339
principle, 341, 344
SPDT (single-pole double throw), 193, 201, 212, 213, 216, 228, 232, 235, 236, 290
Speed controller synthesis, 685–687
block schematic of, 685
Square root, 711
reciprocal of, 713
Stability robustness, 681, 682
Standards
AT&T, 837
BELL, 837
BELLCORE, 837
British Telecom, 837
CNET, 837
HRD, 837
MIL-217 of US Military, 837
most popular, 837
Standby redundancy, 797–799
analysis of, 797
multiple-survivor configuration, 798
with imperfect switching, 799
with multiple failures, 799
Star load, 312
State equation
discretizing, 521, 533–534
extraction, 458, 463–467
State space form, 484
State space method, 542
Steady state
accuracy, 536–538, 549
error, 538
Kalman filter, 667–668
Step-down converter, 205
Step-up converter, 206, 208, 211, 214, 222
inductor current ripple of a, 222
inductor current waveform of a, 222
schematic of a, 219
waveforms of various signals of a, 209, 215
with non-idealities, 213, 219
with power semiconductor switches, 216
Step-up/down converter, 213, 217
topology, 213
Stimuli and response, 806, 807
of control units, 808–809
of subsystems, 808
Switch
single-pole double throw mechanical, 193, 213
SPDT, 201, 201, 212, 216
Switched junction, 468
application of, 470–471
Switches
bidirectional, 42
high-power, 55
ideal, 2
mathematical representation of, 2
semiconductor, 210, 216, 228
switches, 204
uncontrolled, 3
Switching network, 791
dormant failure mode, 792
immediate failure mode, 792

functions performed by, 791
Surge limiting, 117–120
System
complex mechanical, 442
equilibrium model of a, 427
input to a, 425
large-signal model of a, 427
linear time invariant, 542
mesh, 786–788
parallel, 778–786
redundant, 779, 787–799
portraying, 453
response of a, 425
series, 777–778
small-signal model of a, 427
speed of response of a, 538
steady-state model of a, 427
System design
cost, 769
good, 770
life, 769
optimal compromise, 770
performance, 769
primary constraints in, 769–770

T

Tacho, 807
Taylor's series, 426
Technique
agree apportionment, 814
ARINC apportionment, 815
equal apportionment, 814
part count, 837–839
part stress, 839–842
prediction by function, 836–837
similar complexity, 836
similar equipment, 836
Temperature
effect, 173
influence of, 179
junction, 735
parameters depending on, 179
Temperature-regulation coefficient, 172
Tesla, 300
Theorem
Pythagoras, 714
Thermal coefficient, 738, 739
Thermal conductivity of materials, 739
Thermal constants, 739
Thermal resistance of varying cross sections, 759–760
Thermal resistivity, 738, 739
Thermal sense, 807, 809
Thermal stress rating, 221
Thermal viewpoint, 9–11
Thevenin theorem, 167, 174, 179
Thevenin voltage, 179
Time delay, 635
Time update, 594, 665,
in Kalman filter, 667
state-error covariance, 666
value, 666
Torque, 434, 435
drive, 490
equation, 489–490
Tracker, 585–587, 599–600
block diagram of a, 599
control system of a, 631
linear quadratic, 672–674
Transfer function, 435, 510, 511
by hold equivalence, 531
discrete, 531
fourth-order, 446
Transformation
factor, 478
stator to synchronous reference frame, 630
three-phase to two-phase, 630
Transformer, 369
characteristics of an ideal, 388
class, 455
current, 64, 398–403
ideal, 230, 388
impedance, 455
magnetic, 230
potential, 383, 398
practical, 230
secondary, 383
Transformer, secondary
single, 383
multiple, 383
Transistors, 1, 174
bidirectional triode, 38
bipolar junction, 1, 11
bipolar power, 11
calculation of power dissipation of, 1
downstream switching, 128
dynamics characteristics of, 14
gate turn-OFF, 44
insulated gate bipolar, 1, 30
MOSFET (metal oxide semiconductor field effect), 1, 193
NPN, 11
output characteristics of, 14
Paralleling of, 20
PNP, 11
static characteristics of, 13
turn-OFF of, 15
turn-ON of, 16
Transmittance, 755–756
Thyristors, 1
bidirectional, 43
family, 37
GTOs (gate turn-OFF), 1
MOS-controlled, 46
parasitic, 31
Trust factor, 658
Tungsten, 754
Turn ratio, 385, 386

U

Uncertainties
Additive, 681
modeling, 681
multiplicative, 681, 682
Utility power factor converter, 610–613
block diagram, 611

V

Variable
  across, 427
  effort, 427, 454
  flow, 427, 454
  scalar, 447
  state, 429
  through, 427
Vector
  control, 622–631
  input, 485, 486, 488
  output, 485, 486, 488
  state, 485, 486, 488
Voltage
  AC, 95
  bidirectional, 199
  controlling, 600–607
  DC, 95
  forward break-over, 39
  half-wave control output, 136
  handling capacity, 230, 231
  input DC, 149
  input source, 611
  load, 108
  minimum capacitor, 109
  minimum input, 109
  output DC, 149
  pole, 312
  protection, 164
  reverse, 31
  ripple, 109, 124, 128
Voltage regulation
  output, 175
Voltage regulator
  negative, 156–158
  negative output, 158
  negative voltage, 158
Voltage, stator, 484
  direct axis, 481
  quadrature, 481
Volt-second balance, 200, 204, 217, 265, 374
  principle of, 374
Volt-second product, 374

W

Waveform, 228
  capacitor current, 206, 218
  flux, 230
  for a push-pull circuit, 241
  half-wave rectifier, 10
  forward converter, 229
  inductor current, 222
  of various signals of a buck converter, 203, 209
  output voltage, 134
  PWM, 721
  quasi-square, 350
  rectifier output voltage, 138
  six-step, 315
  triangular, 232, 236, 720
Weighted factor, 658
Weighted least squares, 658–659
Weight feedback, 807
Winding
  primary, 383
  secondary, 383
Window utilization factor, 392
Wood, 754

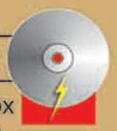
Z

ZCS (zero-current switching), 259
Zener, 152
  breakdown resistance, 177, 183
  breakdown voltage, 153
  current, 151–152, 183
  diode, 36, 151, 152, 166, 174, 180, 183
  ideal, 151
  model, 183
  non-ideal, 153
  voltage reference, 157
Zener diode, 36, 151, 153, 163, 166, 174, 180, 183, 570
  ideal static characteristics of a, 152
  piecewise linear static characteristics of a, 152
  static characteristics of, 153
Zero-current ripple, 261
  buck converter with, 259
operative mode of, 267
Zeros, 436
Zero vector, 347
  choice of, 348
  periods, 347
ZVS (zero-voltage switching), 259
  applied to buck converter, 271–275
  operative modes of, 272
z-Transform, 502–508

About the Book

This book on power electronics spans a wide knowledge base such as power devices, drives, circuit topologies, magnetics, system modeling, control configurations, digital processing, thermal and reliability aspects. The book has been broadly divided into two types of topics viz. (a) circuit-oriented aspects and (b) system-oriented aspects. The first seven chapters deal with circuit-oriented aspects of power electronics systems and the remaining chapters deal with system-oriented aspects like controls and reliability.

Companion CD contains:



- Reliability prediction toolbox
- Bond graph simulation toolbox
- Several circuit and design examples

Salient Features of the Book

Some distinctly unique features of this book are:

- Power semiconductor devices are viewed from the physics, circuit, modeling and thermal viewpoints for a better understanding of the devices.
- AC–DC, DC–DC, DC–AC converters and magnetic devices are treated from both the conceptual and design perspectives.
- A separate chapter is included that addresses the analysis and design of linear regulators.
- A chapter is included to address the modeling methods to obtain dynamic models of power electronics systems. The method of bond graph is introduced for modeling power electronics systems.
- The design of discrete domain controllers in both classical and state space approach are included which addresses the needs of power electronic systems.
- Optimal and robust control design methods as applied to power electronics systems are addressed.
- Discrete numerical algorithms for digital implementation with respect to power electronics systems are addressed in a separate chapter.
- A separate chapter is devoted to the thermal aspects like heat sink sizing for power electronics systems.
- Design integration by specifying and designing for reliability with power electronics system examples is another unique feature of this book.
- The appendices include the following:
  - Derivation of the area product for a saturable-core transformer.
  - Representative list of commonly used core types and their physical parameters.
  - Representative list of commonly used wire gauges.
  - Laplace transforms and *z*-transforms of few time domain signals.
  - List of specifications for the induction motor used for controller design.
  - Description of all the object parameters for various electronic components.
- **Pedagogy includes:**
  - 600+ illustrations and line diagrams.
  - 480+ descriptive questions.
  - 440+ objective questions.
  - 200+ unsolved problems.
  - 50+ explanatory examples and solved problems.

READER LEVEL
Undergraduate

SHELVING CATEGORY
Engineering

Wiley India Pvt. Ltd.
4435-36/7, Ansari Road, Daryaganj,
New Delhi 110 002
Tel: 91-11-43630000
Fax: 91-11-23275895
E-mail: csupport@wileyindia.com
Website: www.wileyindia.com

WILEY

ISBN: 978-81-265-1945-3